



(72) BURRELL, Paul Christopher, AU

(72) BLACKALL, Linda Louise, AU

(72) KELLER, Jurg, AU

(71) CRC FOR WASTE MANAGEMENT AND POLLUTION CONTROL  
LIMITED, AU

(51) Int.Cl.<sup>6</sup> C12N 1/20, C12Q 1/68, C02F 3/34, C07H 21/04

(54) **MICROORGANISMES OXYDANT LES NITRITES DANS L'EAU**

(54) **AQUATIC NITRITE OXIDISING MICROORGANISMS**

(57) The invention relates to the nitrification of wastewater and identification of microorganisms capable of participating in this process. Specifically, the invention provides a consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the Nitrospira phylum. The invention also provides oligonucleotide primers and probes for the amplification or detection of DNA, kits comprising the primers and probes, and methods of detection and quantitating species in a sample.

## ABSTRACT

The invention relates to the nitrification of wastewater and identification of microorganisms capable of participating in this process. Specifically, the invention provides a consortium of  
5 microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum. The invention also provides oligonucleotide primers and probes for the amplification or detection of *Nitrospira* DNA, kits comprising the primers and probes, and methods of detection and quantitating *Nitrospira* species in a sample.

## AQUATIC NITRITE OXIDISING MICROORGANISMS

## TECHNICAL FIELD

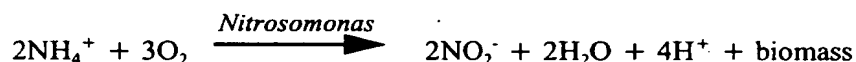
This invention relates to the removal of nitrogenous compounds from wastewater. In particular, the invention relates to an isolated consortium of microorganisms capable of nitrification of wastewater.

5 The invention also relates to methods of identifying microorganisms capable of nitrification of wastewater and oligonucleotide primers and DNA probes suitable for use in the methods.

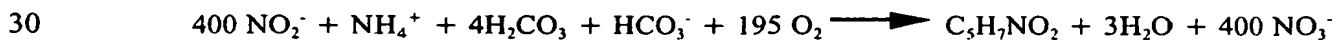
## INTRODUCTION

The removal of nitrogenous compounds from sewage effluents is an important aspect in the remediation of wastewaters. The presence of ammonia, nitrite and nitrate in wastewater discharges  
10 can cause numerous problems ranging from eutrophication (Meganck and Faup, 1988) of the receiving aquatic environment to aspects of public health concern such as nitrate contamination of drinking water. Nitrogen is biologically removed from wastewaters in a two step process of nitrification (ammonia oxidised to nitrate) (Randall, 1992; Robertson and Kuenen, 1991) and denitrification (nitrate reduced to dinitrogen gas that dissipates into the atmosphere) (Blackburn, 1983;  
15 Robertson and Kuenen, 1991). Nitrification is the first and most sensitive step of the process and can be further subdivided into two steps: ammonia oxidation to nitrite and nitrite oxidation to nitrate. The two steps are carried out by separate bacterial groups and for both groups, the total diversity of organisms with this phenotype is small.

Therefore, nitrification is a process where reduced nitrogen compounds, generally ammonium  
20 ( $\text{NH}_4^+$ ), are microbiologically oxidised to nitrate ( $\text{NO}_3^-$ ) via nitrite ( $\text{NO}_2^-$ ) under aerobic conditions (Halling-Sørensen and Jørgensen, 1993). The overall reactions and possible organisms responsible are:



25 The Gram negative chemoautotrophic nitrite oxidising bacteria are physiologically distinct, as they all possess the ability to use nitrite as their energy source and to assimilate  $\text{CO}_2$ , via the Calvin Benson cycle, as a carbon source for cell growth (Bock *et al.*, 1992). For each molecule of  $\text{CO}_2$  fixed, 100 molecules of nitrite need to be oxidized, emphasising the high energy demands placed on these cells. The overall stoichiometry of nitrite oxidation is (Halling-Sørensen and Jørgensen, 1993):



These bacteria can typically also use nitric oxide ( $\text{NO}$ ) instead of  $\text{NO}_2^-$  as an electron source (Bock *et al.*, 1992). Not all of the known nitrifying bacteria are obligate chemoautotrophs. In fact, many strains of *Nitrobacter* can grow well as heterotrophs, where both energy and carbon are obtained from organic carbon sources, or mixotrophically (a combination of both autotrophic and

heterotrophic behaviour). These bacteria are collectively known as facultative chemoautotrophs. Therefore, bacterial strains can grow three ways; aerobically and autotrophically, aerobically and mixotrophically or anaerobically and heterotrophically. In mixotrophic growth,  $\text{NO}_2^-$  is oxidized in preference to organic carbon substrates like acetate, pyruvate and glycerol. Both autotrophic and heterotrophic growth is usually slow and inefficient.

As a generalisation, most strains of *Nitrobacter* seem to be able to grow faster as mixotrophs than as heterotrophs and faster heterotrophically or chemo-heterotrophically than chemoautotrophically.

Four genera are currently recognised: *Nitrobacter*, *Nitrospina*, *Nitrococcus* and *Nitrospira* (Halling-Sørensen and Jørgensen, 1993). *Nitrospina* and *Nitrococcus* are unable to grow heterotrophically or mixotrophically (Bock *et al.*, 1992). One species of *Nitrospira*, *Nitrospira marina*, can grow autotrophically and mixotrophically, (Bock *et al.*, 1992) whereas *Nitrospira moscoviensis* is an obligate autotroph (Ehrich, *et al.*, 1995). These nitrite oxidizers have also been conventionally classified based on phenotypic characters like their cell shape and the ultrastructure of their intracytoplasmic membranes. Doubling times of *Nitrobacter* can range from 12 to 59 hours, or even as long as 140 hours (Halling-Sørensen and Jørgensen, 1993). These are therefore very slow growing bacteria.

In wastewater treatment systems, *Nitrosomonas* (an ammonia oxidizer) and *Nitrobacter* (a nitrite oxidizer) are the two autotrophs presumed to be responsible for nitrification because they are the commonest ammonia and nitrite oxidizers isolated from these environments (Halling-Sørensen and Jørgensen, 1993). Although ammonia oxidizers have been intensively studied by the use of molecular methods (Wagner *et al.*, 1995; Wagner *et al.*, 1996), the nitrite oxidizers have not been similarly investigated. Since the microorganisms responsible for nitrite oxidation in wastewater treatment plants were presumed to be from the genus *Nitrobacter*, mathematical modeling of the process has used data relevant to this genus. However, fluorescent *in situ* hybridization (FISH) probing of activated sludge mixed liquors with *Nitrobacter* specific probes (Wagner *et al.*, 1996) could not confirm the presence of these organisms suggesting that they were not responsible for this major component of nitrogen remediation. Indeed, *Nitrobacter* could not be found in other aquatic environments (Hovanec and DeLong, 1996) when specific FISH probes were employed. It was speculated that other bacteria were likely responsible for nitrite oxidation (Hovanec and DeLong, 1996; Wagner *et al.*, 1996).

Knowledge of the microorganisms responsible for nitrification of wastewater is desirable for the efficient management of treatment systems. It would also be advantageous to have available biomass which can be added to a system to implement or improve nitrification. However, as indicated above, there is no certainty in the art as to the actual microorganisms responsible for nitrification nor are there methods available for identifying such organisms.

## SUMMARY OF THE INVENTION

It is an object of the invention to provide a consortium of microorganisms that can be used for nitrification of wastewater.

A further object of the invention is to provide a method of identifying microorganisms capable  
5 of nitrification of wastewater.

According to a first embodiment of the invention, there is provided a consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum.

According to a second embodiment of the invention, there is provided an oligonucleotide  
10 primer for PCR amplification of *Nitrospira* DNA, said primer comprising at least 12 nucleotides having a sequence selected from:

- (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or
- (ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ  
ID NO: 13.

15 According to a third embodiment of the invention, there is provided a primer pair for PCR amplification of *Nitrospira* DNA, said primer pair comprising:

(a) a first oligonucleotide of at least 12 nucleotides having a sequence selected from one strand of a bacterial 16S rDNA gene; and

(b) a second oligonucleotide of at least 12 nucleotides having a sequence selected from the  
20 other strand of said 16S rDNA gene downstream of said first oligonucleotide sequence; wherein at least one of said first and second oligonucleotides is selected from:

- (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or
- (ii) a DNA sequence having at least 92% identity with any one SEQ ID NO: 1 to SEQ ID  
NO: 13.

25 According to a fourth embodiment of the invention, there is provided a probe for detecting *Nitrospira* DNA, said probe comprising at least 12 nucleotides having a sequence selected from:

- (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or
- (ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ  
ID NO: 13.

30 According to a fifth embodiment of the invention, there is provided a kit comprising:  
at least one primer according to the second embodiment;  
at least one primer pair according to the third embodiment; or  
at least one probe according to the fourth embodiment.

According to a sixth embodiment of the invention, there is provided a method of detecting a  
35 *Nitrospira* species in a sample, said method comprising the steps of:

- (a) lysing cells in said sample to release genomic DNA;
- (b) contacting denatured genomic DNA from step (a) with a primer pair according to the third embodiment;
- (c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and
- (d) detecting said amplification product.

According to a seventh embodiment of the invention, there is provided a method of quantitating the level of a *Nitrospira* species in a sample, said method comprising the steps of:

- (a) lysing cells in said sample to release genomic DNA;
- (b) contacting denatured genomic DNA from step (a) with a primer pair according to the third embodiment;
- (c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and
- (d) detecting said amplification product and quantitating the level of said product by comparison with at least one reference standard.

According to an eighth embodiment of the invention, there is provided a method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:

- (a) lysing cells in said sample to release genomic DNA;
- (b) contacting denatured genomic DNA from step (a) with a labeled probe according to the fourth embodiment under conditions which allow hybridisation of said genomic DNA said probe;
- (c) separating hybridised labeled probe and genomic DNA from unhybridised labeled probe; and
- (d) detecting said labeled probe-genomic DNA hybrid.

According to a ninth embodiment of the invention, there is provided a method of detecting cells of a *Nitrospira* species in a sample, said method comprising the steps of:

- (a) treating cells in said sample to fix cellular contents;
- (b) contacting said fixed cells from step (a) with a labeled probe according to the fourth embodiment under conditions which allow said probe to hybridise with RNA within said fixed cell;
- (c) removing unhybridised probe from said fixed cells; and
- (d) detecting said labeled probe-RNA hybrid.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing influent and effluent NO<sub>2</sub>-N concentrations for an automated laboratory-scale reactor operating as a sequencing batch reactor at 2 cycles/day with strong selection for nitrite oxidising biomass (NOSBR).

Figure 2 is a graph showing influent and effluent NO<sub>2</sub>-N concentrations of the NOSBR operating at 4 cycles/day.

Figure 3 is a graph of mixed liquor nitrite-N concentrations during the react period of the NOSBR cycle for attached growth and for suspended growth.

5 Figure 4 is a graph showing nitrite-N and nitrate-N concentrations in the mixed liquor during the react period of the NOSBR.

Figure 5 is a graph showing mixed liquor nitrite-N concentrations during the react period in three stages of the NOSBR operated at 2 cycles/day with different concentrations of nitrite in the feed.

10 Figure 6 is a graph of mixed liquor nitrite-N concentrations during the react period in three representative cycles during operation of the NOSBR at 4 cycles/day.

Figure 7 is an evolutionary distance tree derived from a comparison of 16S rDNA sequences from nitrite oxidising bacteria and clone sequences from three different 16S rDNA clone libraries (RC, GC, and SBR).

15 Figure 8 is an alignment of sequences of 16S rDNA from *Nitrospira* clones identified in a nitrite-oxidising SBR and from other sources.

Figure 9 depicts the results of agarose gel electrophoresis of PCR-amplified DNA using genomic DNA from various *Nitrospira* clones as template.

#### BEST MODE AND OTHER MODES OF CARRYING OUT THE INVENTION

The following abbreviations are used hereafter:

20	SBR	sequencing batch reactor
	NOSBR	nitrite oxidising SBR
	NOM	nitrite oxidising medium
	HRT	hydraulic retention time
	MLSS	mixed liquor suspended solids
25	BNR	biological nutrient removal
	DO	dissolved oxygen
	PCR	polymerase chain reaction
	REA	restriction enzyme analysis
	OTU	operational taxonomic unit
30	bp(s)	base pair(s)

The one-letter code for nucleotides in DNA conforms to the IUPAC-IUB standard described in *Biochemical Journal* 219, 345-373 (1984).

The term "comprise", or variations of the term such as "comprises" or "comprising", are used herein to denote the inclusion of a stated integer or stated integers but not to exclude any other

integer or any other integers, unless in the context or usage an exclusive interpretation of the terms is required.

The present inventors have developed a specific nitrifying biomass that is largely comprised of bacteria that are most closely related to *Nitrospira moscoviensis*. It is believed that a range of species of *Nitrospira* are involved in the process. The inventors have shown that these bacteria are likely to be more dominant in reactors with good nitrification performance than bacteria from the genus *Nitrobacter*. A range of studies have failed to find *Nitrobacter* in nitrifying processes (Hovanec & DeLong, 1996; Wagner *et al.*, 1996) and evidence is provided below that the organisms responsible for this important biochemical reaction in wastewater treatment processes (both suspended and attached growth processes) are from the *Nitrospira* phylum in the domain *Bacteria*.

With reference to the first embodiment of the invention, the nitrifying biomass can be produced by presenting a feed comprising nitrite, dissolved oxygen and dissolved carbon dioxide but which is free of organic carbon to seed sludge from any sewage plant exhibiting nitrification. The seed sludge is advantageously from a domestic wastewater treatment plant but can also be from an abattoir wastewater treatment plant. The nitrite component of the feed can be as low as about 400 mg/L nitrite-N. The oxygen and carbon dioxide can conveniently be provided as air bubbled through the solution.

Turning to the second embodiment of the invention, oligonucleotide primers typically have a length of about 12 to 50 nucleotides. A preferred length is 12 to 22 nucleotides. Particularly preferred primers are the following:

5' CGGGAGGGAAGATGGAGC 3' (SEQ ID NO: 14)

5' CCAACCCGGAAAGCGCAGAG 3' (SEQ ID NO: 15)

5' AGCCTGGCAGTACCCTCT 3' (SEQ ID NO: 16)

Oligonucleotide primer pairs according to the third embodiment of the invention comprise an oligonucleotide primer that will anneal to one strand of the target sequence and a second oligonucleotide primer which will anneal to the other, complementary, strand of the target sequence. It will be appreciated that the second oligonucleotide primer must anneal to the complementary strand downstream of the first oligonucleotide primer sequence, which occurs in the complementary strand, to yield a double stranded amplification product in the PCR. The amplification product is of a size that facilitates detection. Typically, the first and second oligonucleotide primer sites in the target DNA are separated by 50 to 1,400 bps. A preferred separation is 400 to 1,000 bps.

The probes of the fourth embodiment, as indicated above, can have a size as small as 12 nucleotides. Typically, however, probes have a length of 15 to 50 nucleotides. A preferred probe length is 15 to 22 nucleotides, particularly for *in situ* hybridisation according to the method of the ninth embodiment.



The oligonucleotide primers included in kits according to the fifth embodiment of the invention can be individual oligonucleotide primers appropriate for the detection of *Nitrospira* or a primer pair. Oligonucleotide primer pairs are advantageously provided as compositions. Additional oligonucleotide primers can also be included in kits for use in control reactions. For detection purposes, DNA probes  
5 can also be included in kits.

Kits according to the fifth embodiment of the invention can further comprise reagents used in PCR and hybridisation reactions. Such reagents include buffers, salts, detergents, nucleotides and thermostable polymerase. Such reagents are advantageously provided as solutions to facilitate execution of PCR or hybridisation. Solutions can be compositions comprising a number of reagents as is well  
10 known in the art.

The general techniques used in the methods of the sixth to ninth embodiments, and factors to be considered in selecting PCR primers and probes, will be known to those of skill in the art. Such techniques are described, for example, in Sambrook *et al.* (1989) and Stackebrandt and Goodfellow (1991), the entire contents of which are incorporated herein by cross reference. Particularly relevant  
15 chapters in Stackebrandt and Goodfellow are Chapter 7, "The Polymerase Chain Reaction" by S. Giovannoni, and Chapter 8, "Development and Application of Nucleic Acid Probes" by D. A. Stahl and R. Amann.

Non-limiting examples of the invention will now be provided.

#### General Methods

20 The total community DNAs from the NOSBR sludge (RC) and the seed sludge (GC) were isolated, the 16S rDNAs were polymerase chain reaction (PCR) amplified and cloned using previously published methods (Blackall, 1994; Blackall *et al.*, 1994; Bond *et al.*, 1995). Inserts from 102 clones in the RC library were amplified and grouped by *Hae*III restriction enzyme digestion banding profiles (REA) into operational taxonomic units (OTUs) (Weidner *et al.*, 1996). Clone inserts from  
25 representatives of RC OTUs and all 77 clones from the GC library were PCR amplified and partially sequenced (Blackall, 1994) using 530f (Lane, 1991) primer. Inserts from a selection of clones were fully sequenced (Blackall, 1994). Sequence data were analysed according to previously published methods (Blackall *et al.*, 1994) which included BLAST (Altschul *et al.*, 1990) comparisons and phylogenetic analyses (Felsenstein, 1993).

30

#### Example 1

##### Selection of a Nitrifying Biomass

In this example, we describe the use of a laboratory-scale reactor as a sequencing batch reactor (SBR) with strong selection for a nitrite oxidising biomass. Seed sludge was from the Merrimac domestic wastewater treatment plant operated by the Gold Coast City Council and located

at Merrimac, Queensland 4226, Australia. The reactor set-up will be hereafter referred to as the "Nitrite Oxidising SBR", or "NOSBR".

*Reactor.* A laboratory chemostat with a working volume of 1 L was operated in the dark at 24°C as the NOSBR. The influent nitrite oxidising medium (NOM) was a synthetic waste water mix comprising per L: 400 to 3,200 mg KNO<sub>2</sub>, 3.75 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 250 mg CaCl<sub>2</sub>·2H<sub>2</sub>O, 10 g KH<sub>2</sub>PO<sub>4</sub>, 10 g K<sub>2</sub>HPO<sub>4</sub>, 200 mg FeSO<sub>4</sub>·7H<sub>2</sub>O, and 20 g NaHCO<sub>3</sub>. The pH of the medium was adjusted to 7.0, but the reactor was not equipped with pH control. Dissolved oxygen was maintained at 1.6-2.0 mg/L and CO<sub>2</sub> was introduced by bubbling air through the liquid in the NOSBR. Surface biomass growth was precluded by regular scrubbing of all solid surfaces with a brush. Four cycles per day giving a hydraulic retention time (HRT) of 12 hr were performed with the following sequences:-

- 1) Feed of 500 ml of fresh medium - 30 min (0 to 0.5 hr)
- 2) React (aeration) - 4.5 hr (0.5 to 5 hr)
- 3) Settle - 40 min (5 to 5.7 hr)
- 4) Decant 500 ml of supernatant - 20 min (5.7 to 6 hr)
- 5) Total time per cycle - 6 hr.

Automatic timers controlled the magnetic stirrer (100 rpm), peristaltic pumps (feed and decant), and air pump for the cycles. Sludge biomass was not wasted from the reactor, but periodically, biomass was collected for testing which facilitated maintenance of a relatively steady amount of biomass in the SBR.

At start up, 1 L of mixed liquor suspended solids (MLSS) from a full scale Biological Nutrient Removal (BNR, nitrogen and phosphorus removal) plant was added to the NOSBR which was operated manually with the NOM. Initial manual and then automatic operation with 2-cycles per day (feed - [500 ml] 40 min; react - 10 hr; settle - 40 min; and decant [500 ml] - 40 min) occurred for some months before initiation of the 4-cycles per day scheme (see above).

*Monitoring.* Chemical analyses of feed, mixed liquor and effluent were regularly done for nitrite-N (NO<sub>2</sub>-N), nitrate-N (NO<sub>3</sub>-N), and ammonium-N (NH<sub>4</sub><sup>+</sup>-N) using spectrometric assays (Merck, Melbourne, Australia). To preclude the removal of excessive biomass, these analyses were done with 2 ml samples. The MLSS of the NOSBR was determined in duplicate 10 ml samples of mixed liquor. These were filtered onto pre-dried Whatman GF/C filters, and then dried to a constant weight at 105 degree C. A pH meter was used to periodically monitor pH in the mixed liquor and effluent. A portable dissolved oxygen (DO) meter and probe were used to periodically monitor the DO in the NOSBR.

*Results of operation.* Varying influent nitrite levels were employed to study a range of features of the selected nitrite oxidising biomass. The operating data for the influent and effluent nitrite levels

of the NOSBR during the automated 2 cycles/day period are presented in Figure 1 and for the automated 4 cycles/day in Figure 2. The data presented in these figures show that the microbial community are able to remove all the nitrite from the influent in a matter of hours.

#### Attributes of the NOSBR mixed liquor

5       1. *Suspended versus attached growth - 2 cycles/day.* To generate attached growth, the regular scrubbing regime of the reactor was suspended for two weeks. The vast bulk of the biomass was then attached to surfaces in the reactor. The little remaining suspended biomass was discharged from the reactor which was then filled with 1 L of half strength NOM. Regular sampling and nitrite analyses were done during the react period of one cycle with all the biomass attached to the reactor surfaces.  
10       The results of this experiment are presented in Figure 3. The results show that suspended biomass has twice the nitrite oxidation rate than the attached biomass but both systems are effective in removing nitrite from the influent.

Following the experiment described in the previous paragraph, the biomass was completely scrubbed from the surfaces to the liquid. The reactor was operated for two cycles with biomass  
15       scrubbing. A similar one-cycle study was performed as with the attached growth but with all biomass suspended. The biofilm growth exhibited a nitrite oxidation rate of 29 mg NO<sub>2</sub>-N/hr and the suspended growth form showed a rate of 58 mg NO<sub>2</sub>-N/hr. It was assumed that the biomass concentration was the same for both studies since none had been removed between them.

20       2. *pH correlation with nitrification.* It was observed that when the pH of the effluent fell below 7.4, nitrite-N was present in the effluent. If the pH rose above 7.4 for short periods, no effect to nitrification was observed. Therefore, pH values below 7.4 were detrimental to nitrification.

25       3. *Cyclic studies.* Figure 4 shows the results for periodic measurements of nitrite-N and nitrate-N during the react period of the reactor during 2 cycles/day. The results presented in these figures show that the bacterial population in the reactor oxidised nitrite to nitrate in a stoichiometric manner with 160 mg/l of nitrite-N being oxidised to 160 mg/l of nitrate-N (170 mg/l at the start of the react period and 330 mg/l when the nitrite-N was exhausted). The rate of nitrite oxidation and nitrate production also appeared to be linear, showing that the oxidation process was not limited by any external factors.

30       Studies measuring nitrite reaction in the reactor are shown for both 2 cycles/day (Figure 5) and 4 cycles/day operation (Figure 6). The significance of these results is that the biomass is robust in its capacity to oxidise nitrite under a range of operating conditions.

#### Example 2

##### The Microbiology of the NOSBR

35       In this example, we describe the microbiological characterisation of the nitrifying microorganisms present in the biomass selected in the NOSBR described in Example 1. Methods used

in the characterisation have been described by Blackall (1994) and Bond *et al.* (1995), the entire contents of which disclosures are incorporated herein by cross-reference.

Total microbial community DNA from both the seed BNR sludge (GC) and from the reactor after six months of operation (RC) was obtained. The 16S rDNA from each DNA extract were separately amplified by polymerase chain reaction (PCR), and then for each, clone libraries were prepared (Blackall, 1994; Bond *et al.*, 1995).

Inserts from a total of 77 clones from the GC clone library were partially sequenced with the primer 530f and phylogenetically analysed (Blackall *et al.*, 1994) (Table 1). The majority of the clone sequences grouped with the proteobacterial phylum, while 4% (3 clones; GC3, GC86 and GC109) grouped with the phylum *Nitrospira*.

Table 1

Phyla from the Domain Bacteria Represented in the GC Clone Library

Phylum in Domain Bacteria	Percentage in clone library
Proteobacteria	
Alpha	5
Beta	29
gamma	18
delta	4
High mol%G+C Gram positives	10
Low mol%G+C Gram positives	7
<i>Flexibacter/Cytophaga/Bacteroides</i>	5
<i>Nitrospira</i>	4
Planctomycetales	9
Unaffiliated	9

Restriction Enzyme Analysis (REA) of the RC library was done to group clones into operational taxonomic units (OTUs) in advance of partial or complete clone insert sequencing (Weidner *et al.*, 1996). Thirteen different OTUs were found when *HaeIII* was employed as the restriction enzyme to digest the inserts from 102 clones. The large majority of the clone inserts (88% or 90 clones) were found in one OTU while the remaining 12% (12 clones) comprised individuals in 12 other OTUs. Each of the clone inserts from the latter 12 OTUs and six of the large former group (RC7, RC11, RC16, RC25, RC73, and RC99) were partially sequenced and phylogenetically analysed. These six and one of the other OTUs (RC90) were found to have partial insert sequences that phylogenetically grouped with the *Nitrospira* phylum. From this analysis, it was concluded that 91 clones or 89% of the clone library originated from bacteria in the *Nitrospira* phylum. In the

phylogenetic analysis, one of the other OTUs (RC44) grouped with *Nitrobacter*. It was concluded that the organisms responsible for nitrification in the NOSBR were likely to be from the *Nitrospira* phylum.

Near complete insert sequence analyses were done for the following clones:

- 5 - six RC clones of the original partial sequences - RC7, RC11, RC25, RC73, RC90, and RC99 (RC16 omitted);
- two RC clones from the *Nitrospira* OTU (RC14 and RC19);
- one of the three GC *Nitrospira* clones (GC86); and
- 10 - four clones from a clone library prepared by Bond *et al.* (1995) that phylogenetically grouped in the *Nitrospira* phylum.

The data were phylogenetically analysed as shown in Figure 7. The two clone clades would likely comprise two separate species with the RC clones possibly comprising more than one species.

Sequences of clones from the two *Nitrospira* clades were subjected to direct pairwise sequence comparison. The results of this comparison are presented in Table 2. The table is a similarity matrix showing the percent similarity between 16S rDNA sequences of *Nitrospira moscoviensis*, *Nitrospira marina* and 13 near complete sequences from clone inserts from a full scale biological nutrient removal activated sludge plant (GC86), from the NOSBR (RC clone numbers) and from clones for which the partial sequences had been previously reported (SBR clones; Bond *et al.*, 1995). The similarity matrix showed that the first clade (SBR1015, SBR1024, SBR2046, GC86) had an average 15 16S rDNA comparison value of 99.4% while for the second clade (RC7, RC11, RC14, RC19, RC25, RC73, RC90, RC99, SBR2016), this value was 98.7%. The highest comparative value between an RC clone sequence and *N. moscoviensis* was 93.4% for RC25. From the sequence data analysis, the two clone clades would likely comprise two separate species, with the RC clones possibly comprising more than one species.

25 Sequence data for the SBR, GC and RC clones are presented in Figure 8. In this figure, sequences are divided into blocks with numbers given in square brackets above each block. The clone identification is given at the left of a line of sequence in each block. Dashes represent unknown nucleotides while full stops represent alignment breaks.

The sequences of clones are also presented as sequence listings as follows:

<u>Clone</u>	<u>Sequence Listing Number</u>
SBR1024	1
SBR1015	2
GC86	3
SBR2046	4
RC25	5
RC19	6
SBR2016	7
RC7	8
RC14	9
RC99	10
RC11	11
RC73	12
RC90	13

Table 2

Species or clone	Percent sequence similarity with species of strain number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. <i>Nitrospira moscoviensis</i>															
2. SBR1024	96.3														
3. SBR1015	96.1	99.6													
4. GC86	96.1	99.6	99.4												
5. SBR2046	95.8	99.3	99.4	99.2											
6. RC25	93.4	93.4	93.6	93.6	93.1										
7. RC19	93.2	93.1	93.0	93.2	92.7	98.8									
8. SBR2016	93.0	92.7	92.8	92.6	92.4	99.1	98.7								
9. RC7	92.9	93.1	93.2	92.9	92.8	98.7	98.7	98.5							
10 RC14	92.8	93.0	93.1	93.1	92.7	98.7	98.9	98.5	99.3						
11 RC99	92.7	92.9	93.0	93.0	92.6	98.5	98.7	98.4	99.2	99.6					
12 RC11	92.6	92.8	93.0	92.9	92.5	98.5	98.7	98.4	99.0	99.5	99.7				
13 RC73	92.2	92.5	92.6	92.6	92.1	98.0	98.2	97.9	98.7	99.1	99.4	99.4			
14 RC90	92.1	92.1	92.3	92.2	91.8	98.1	98.6	98.0	98.1	98.6	98.8	98.8	99.0		
15 <i>Nitrospira marina</i>	88.7	88.2	88.3	88.3	87.8	88.1	87.6	87.2	87.2	87.1	87.1	87.1	86.5	86.6	
16 <i>Nitrospira marina</i>	88.0	88.0	88.2	88.1	87.7	87.9	87.5	87.2	87.2	87.1	87.1	87.1	86.5	86.6	99.9

## Example 3

Identification of *Nitrospira* Species

Primers for use in a diagnostic PCR for the *Nitrospira moscoviensis* clade of Figure 7 (see Example 2) were designed from aligned sequence datasets (see Tables 3-5 below ).

- 5 Table 3 is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS457f (SEQ ID NO: 14) for the *Nitrospira moscoviensis* clade. In the table, mismatches with the primer sequence are in bold type and are underlined. The melting temperature calculated for MOS457f was 60°C and a fragment size of approximately 1052 nucleotides was calculated in a PCR with primer 1492r. The MOS457f  
10 sequence corresponds to the sequence at positions 440 to 457 of the *E. coli* 16S rDNA gene.

Table 3

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS457f primer (SEQ ID NO: 14)	CGGGAGGGAAGATGGAGC	-
<i>Nitrococcus mobilis</i> (SEQ ID NO: 17)	CAGCCGGGAGGAAAAGCA	10
<i>Magnetobacterium bavaricum</i> (SEQ ID NO: 18)	TGTAGGGAAGATGATGA	8
<i>Nitrobacter hamburgensis</i> (SEQ ID NO: 19)	TGTGCGGGAAGATAATGA	7
<i>Nitrospina gracilis</i> (SEQ ID NO: 20)	CGGGTGGGAAGAACA AAA	6
<i>Nitrospira marina</i> (SEQ ID NO: 21)	CATGAGGAAGATAAAGT	6
SBR1015 (SEQ ID NO: 22)	CGGCAGGGAAGATGGAAC	2
SBR1024 (SEQ ID NO: 22)	CGGCAGGGAAGATGGAAC	2
SBR2016 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
SBR2046 (SEQ ID NO: 24)	CCGCAGGGAAGATGGAAC	3
RC7 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC11 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC14 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC19 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC25 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC73 (SEQ ID NO: 25)	CGGGAGGGAAGATGGAAC	1
RC90 (SEQ ID NO: 25)	CGGGAGGGAAGATGGAAC	1
RC99 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC44 ( <i>Nitrobacter</i> clone) (SEQ ID NO: 26)	CGTGC GGGAAGATAATGA	6
GC86 (SEQ ID NO: 27)	CGGCAGGGAAGATGGAAC	2
<i>Nitrospira moscoviensis</i> (SEQ ID NO: 28)	CGGGAGGGAAGATGGACG	2



Like Table 3, Table 4 is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS638f (SEQ ID NO: 15) for the *Nitrospira moscoviensis* clade. Again, mismatches with the primer sequence are in bold and are underlined. The calculated melting temperature for this primer was 66°C and a fragment size of approximately 873 nucleotides was calculated in a PCR with primer 1492r. The MOS638f sequence corresponds to the sequence at positions 619 to 638 of the *E. coli* 16S rDNA gene.

Table 4

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS638f primer (SEQ ID NO: 15)	CCAACCCGGAAAGCGCAGAG	-
<i>Nitrococcus mobilis</i> (SEQ ID NO: 29)	<u>T</u> CAACCTGGGAAT <u>T</u> GCATCC	8
<i>Magnetobacterium bavaricum</i> (SEQ ID NO: 30)	<u>T</u> CAACCCGGGAAT <u>T</u> GCCTTG	7
<i>Nitrobacter hamburgensis</i> (SEQ ID NO: 31)	<u>T</u> CAACTCCAGAACTGCCTTT	11
<i>Nitrospina gracilis</i> (SEQ ID NO: 32)	<u>T</u> CAACCGTGGAAT <u>T</u> GCCTTT	10
<i>Nitrospira marina</i> (SEQ ID NO: 33)	<u>T</u> TAACCGGGAAAGGT <u>C</u> GAGA	9
SBR1015 (SEQ ID NO: 34)	C <u>T</u> AACCCGGAAAGTGC <u>G</u> GAG	3
SBR1024 (SEQ ID NO: 34)	C <u>T</u> AACCCGGAAAGTGC <u>G</u> GAG	3
SBR2016 (SEQ ID NO: 35)	CCAACCCG <u>A</u> AAAGCGCAGAG	1
SBR2046 (SEQ ID NO: 34)	C <u>T</u> AACCCGGAAAGTGC <u>G</u> GAG	3
RC7 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC11 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC14 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC19 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC25 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC73 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC90 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC99 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC44 ( <i>Nitrobacter</i> clone) (SEQ ID NO: 37)	<u>T</u> CAACTCCAGAACTGCCTTT	11
GC86 (SEQ ID NO: 34)	C <u>T</u> AACCCGGAAAGTGC <u>G</u> GAG	3
<i>Nitrospira moscoviensis</i> (SEQ ID NO: 38)	CCAACCCGGAAAGCGCAGAG	0

10 Table 5, is again an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS635r (SEQ ID

NO: 16) for the *Nitrospira moscoviensis* clade. The melting temperature calculated for this primer was 58°C and a fragment size of approximately 625 nucleotides was calculated in a PCR with primer 27f. The MOS635r sequence corresponds to the sequence at positions 635 to 652 of the *E. coli* 16S rDNA sequence.

5

Table 5

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS635r primer (SEQ ID NO: 16)	AGCCTGGCAGTACCCTCT	-
<i>Nitrococcus mobilis</i> (SEQ ID NO: 39)	AGCC <u>AAAC</u> AGTATCGGAT	7
<i>Magnetobacterium bavaricum</i> (SEQ ID NO: 40)	AGTTAAACAGTTTTCAAG	11
<i>Nitrobacter hamburgensis</i> (SEQ ID NO: 41)	AGACCTTCAGTATCAAAG	9
<i>Nitrospina gracilis</i> (SEQ ID NO: 42)	AGCCGAATAGTTTCAAAC	10
<i>Nitrospira marina</i> (SEQ ID NO: 43)	AGCTGAATAGTTCCTCTC	10
SBR1015 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
SBR1024 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
SBR2016 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
SBR2046 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
RC7 (SEQ ID NO: 46)	AGCCTGGCAGTACCCCTCT	1
RC11 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC14 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC19 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC25 (SEQ ID NO: 47)	AGCCTGGCAGTACCCTCT	1
RC73 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC90 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC99 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC44 ( <i>Nitrobacter</i> clone) (SEQ ID NO: 48)	AGATCCTCAGTATCAAAG	10
GC86 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
<i>Nitrospira moscoviensis</i> (SEQ ID NO: 49)	AGCCTGGCAGTACCCTCT	0

The three primers defined above in Tables 3 to 5 were included in separate primer pairs which pairs were then tested in PCR amplifications using genomic DNA from various *Nitrospira* clones as template. The PCRs were carried out according to methods detailed in Sambrook *et al.* (1989) at an annealing temperature of 62°C.

10

The results of electrophoretic analysis of PCRs on an agarose gel are presented in Figure 9. Details of the material analysed in each lane of the gel are given in Table 6. The marker DNA was

*Hae*III-digested  $\phi$ X174 DNA. The sizes of the  $\phi$ X174 fragments are given on the left-hand side of the figure.

Table 6

Lane	Primer pair used	Mismatches between primer and template
1	( <i>Hae</i> III-digested $\phi$ X174 DNA)	
2	MOS457f, 1492r	0 mismatches with MOS457f
3	MOS457f, 1492r	1 mismatch with MOS457f
4	MOS457f, 1492r	2 mismatches with MOS457f
5	( <i>Hae</i> III-digested $\phi$ X174 DNA)	
6	MOS638f, 1492r	0 mismatches with MOS638f
7	MOS638f, 1492r	1 mismatch with MOS638f
8	MOS638f, 1492r	3 mismatches with MOS638f
9	( <i>Hae</i> III-digested $\phi$ X174 DNA)	
10	MOS635r, 27f	0 mismatches with MOS635r
11	MOS635r, 27f	1 mismatch with MOS635r
12	MOS635r, 27f	4 mismatches with MOS635r

5 The results presented in Figure 9 show that an amplicon of the appropriate size was obtained in reactions where there was up to one mismatch between a primer and the template but that no amplicon was produced where there was a greater degree of mismatch.

When the three primer pairs used for the results presented in Figure 9 were used with clone RC44 (closest match to *Nitrobacter*), no amplicons were produced.

10 The primer NIT3 (Wagner *et al.* 1996; SEQ ID NO: 50) was used in a diagnostic PCR for *Nitrobacter*. NIT3 was designed originally for fluorescent *in situ* hybridisation experiments. The specificity of this primer can be appreciated from the sequence alignment presented in Table 7 which is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla against NIT3. A melting temperature of 60°C was calculated for NIT3 and a  
15 fragment size of approximately 1020 nucleotides in a PCR with primer 27f as experimentally determined. The NIT3 sequence corresponds to the sequence at positions 1031 to 1048 of the *E.coli* 16S rDNA gene.

Table 7

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
NIT3 primer (SEQ ID NO: 50)	CCTGTGCTCCATGCTCCG	-
<i>Nitrobacter hamburgensis</i> (SEQ ID NO: 51)	CCTGTGCTCCATGCTCCG	0
<i>Nitrospina gracilis</i> (SEQ ID NO: 52)	CCTGTGCA <u>AAGGGCCCCGA</u>	9
<i>Nitrococcus mobilis</i> (SEQ ID NO: 53)	CCTGT <u>CATCCGGT</u> TCCCG	7
<i>Nitrospira moscoviensis</i> (SEQ ID NO: 54)	CCTG <u>AGCACGCTGGT</u> ATT	8
<i>Nitrospira marina</i> (SEQ ID NO: 55)	CCTG <u>AGCTCGCTCCCCT</u> T	7
<i>Magnetobacterium bavaricum</i> (SEQ ID NO: 56)	CCTGTGCA <u>AAGCTCTCCCT</u>	8
SBR1015 (SEQ ID NO: 57)	CCTG <u>AGCAGGATGGT</u> ATT	8
SBR1024 (SEQ ID NO: 57)	CCTG <u>AGCAGGATGGT</u> ATT	8
SBR2016 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
SBR2046 (SEQ ID NO: 57)	CCTG <u>AGCAGGATGGT</u> ATT	8
RC7 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC11 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC14 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC19 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC25 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC73 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
RC90 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8
GC86 (SEQ ID NO: 59)	CCTG <u>AGCAGGATGGT</u> GTT	8
RC99 (SEQ ID NO: 58)	CCTG <u>AGCACGCTGGT</u> ATT	8

Results of PCRs with the primer pair NIT3 and 27f showed that the NIT3 primer specifically amplified only RC44 clone inserts (*Nitrobacter*) and not those from *Nitrospira* clones.

5

The different primer pairs were then used with DNAs extracted from sludges and the results are tabulated below in Table 8. The scorings presented in the table were generated by quantitating by eye the intensity of the amplificate in a stained gel. A definition of the scoring follows: - = no band; +/- = very faint band; + through + + + + = increasing intensity of the amplificate.

Table 8

Wastewater Treatment Plant	Performance	MOS635r-27f	NIT3-27f
		620 bp	1020 bp
Oxley	Full nitrification	++++	++
Merrimac	Full nitrification	++++	++
Loganholme	Full nitrification	+++	+/-
Gibson Island	Full nitrification	+++	-
Fairfield	No nitrification	+/-	+++
Cannon Hill	Full nitrification	+	+
NOSBR	NO <sub>2</sub> <sup>-</sup> oxidation	+++++	++++
Saline waste water BNR SBR	Partial nitrification	+/-	++
Nitrifying biofilm reactor	Full nitrification	++++	++++
Phenol/cyanide removing SBR	No nitrification	+/-	++
BNR SBR	Full nitrification	+	+

These results show that in plants having good nitrification, *Nitraspira* species were present as evidenced by amplification of target DNA with the selected primer pairs.

## REFERENCES

- Blackall, L. L. (1994). Molecular identification of activated sludge foaming bacteria. *Water Science and Technology* **29-7**, 35-42.
- Blackall, L. L., Seviour, E.M., Cunningham, M.A., Seviour, R.J., and Hugenholtz, P. (1994).  
5 "Microthrix parvicella" is a novel, deep branching member of the actinomycetes subphylum. *Systematic and Applied Microbiology* **17**, 513-518.
- Blackburn, T. H. (1983). The Microbial Nitrogen Cycle. In *Microbial Geochemistry*. (pp. 63-89). Edited by W. E. Krumbein. Oxford: Blackwell Scientific Publications.
- Bock, E., Koops, H., Ahlers, B., and Harms, H. (1992). Oxidation of inorganic nitrogen compounds  
10 as energy source. In *The Prokaryotes - A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, pp. 414-430. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. New York: Springer-Verlag.
- Bond, P.L., Hugenholtz, P., Keller, J., and Blackall, L.L. (1995). Bacterial community structures of  
phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors.  
15 *Applied and Environmental Microbiology*, **61**, 1910-1916.
- Burrell, P., and Blackall, L.L. (in press). The microbiology of nitrogen removal in activated sludge systems. In *The Microbiology of Activated Sludge*, Edited by R. J. Seviour & L. L. Blackall.
- Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W. and Bock, E. (1995). A new obligately  
chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its  
20 phylogenetic relationship. *Archives of Microbiology*, **164**, 16-23.
- Halling-Sørensen, B., and Jørgensen, S.E. (1993). *The Removal of Nitrogen Compounds from Wastewater*. Amsterdam: Elsevier.
- Hovanec, T. A. & DeLong, E. F. (1996). Comparative Analysis of Nitrifying Bacteria Associated with Freshwater and Marine Aquaria. *Applied and Environmental Microbiology*, **62**, 2888-2896.
- 25 Meganck, M.T.J., and Faup, G.M. (1988). Enhanced biological phosphorus removal from waste waters. *Biotreatment Systems*, **3**, 111-204.
- Randall, C. W. (1992). Introduction and Principles of Biological Nutrient Removal. In *Design and Retrofit of Wastewater Treatment Plants for Biological Nutrient Removal*, (pp. 7-84). Edited by C. W. Randall. Lancaster: Technomic Publishing Company Inc.
- 30 Robertson, L. A. & Kuenen, J. G. (1991). Physiology of Nitrifying and Denitrifying Bacteria. In *Microbial Production and Consumption of Greenhouse Gases: Methane, Nitrogen Oxides and Halomethanes*, (pp. 189-199). Edited by J. E. Rogers & W. B. Whitman. Washington D.C.: American Society for Microbiology.
- Sambrook, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Second  
35 Edition, Cold Spring Harbor Laboratory Press, Plainview, New York, 1989.

- Stackebrandt, E., and Goodfellow, M., eds, *Nucleic Acid Techniques in Bacterial Systematics*, John Wiley & Sons, New York, 1991
- Wagner, M., Rath, G., Koops, H.-P., Flood, J., and Amann, R. (1996). In situ analysis of nitrifying bacteria in sewage treatment plants. *Water Science and Technology*, **34**, 237-244.
- 5 Weidner, S., Arnold, W. and Pühler, A. (1996). Diversity of Uncultured Microorganisms Associated with the seagrass *Halophila stipulacea* Estimated by Restriction Fragment Length Polymorphism Analysis of PCR-Amplified 16S rRNA Genes. *Applied and Environmental Microbiology*, **62**, 766-771.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: CRC for Waste Managment and Pollution Control Limited  
(B) STREET: High Street  
(C) CITY: Kensington  
(D) STATE: New South Wales  
(E) COUNTRY: Australia  
(F) POSTAL CODE (ZIP): 2033

(ii) TITLE OF INVENTION: Aquatic Nitrite Oxidising Microorganisms

(iii) NUMBER OF SEQUENCES: 59

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1428 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAAGTCGAGC GAGAAGACGT AGCAATACGT TTGTAAAGCG GCGAACGGGT GAGGAATACA	60
TGGGTAACCT ACCTTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG	120
ACTCCTGGTC TGC GGATCGG GAGAGAAAGC GATACCGTGG GTATCGCGCT CTTGGATGGG	180
CTCATGTCCT ATCAGCTTGT TGGTGAGGTA ACGGCTCACC AAGGCTTCGA CGGGTAGCTG	240
GTCTGAGAGG ACGATCAGCC AACTGGCAC TGCGACACGG GCCAGACTCC TACGGGAGGC	300
AGCAGTAAGG AATATTGCGC AATGGGCGAC AGCCTGACGC AGCNACGCCG CGTGGGGGAT	360



GAAGGTCTTC GGATTGTAAA CCCCTTTCGG CAGGGAAGAT GGAACGGGTA ACCGTTTCGGA 420  
 CGGTACCTGC AGAAGCAGCC ACGGCTAACT TCGTGCCAGC AGCCGCGGTA ATACGAAGGT 480  
 5 GGCAAGCGTT GTTCGGATTT ACTGGGCGTA CAGGGAGCGT AGGCGGTTGG GTAAGCCCTC 540  
 CGTGAAATCT CCGGGCCTAA CCCGGAAAGT GCGGAGGGGA CTGCTCGGCT AGAGGATGGG 600  
 10 AGAGGAGCGC GGAATTCCCG GTGTAGCGGT GAAATGCGTA GAGATCGGGA GGAAGGCCGG 660  
 TGGCGAAGGC GGCCTCTGG AACATTTCTG ACGCTGAGGC TCGAAAGCGT GGGGAGCAAA 720  
 CAGGATTAGA TACCCTGGTA GTCCACGCCT TAAACGATGG ATACTAAGTG TCGGCGGGTT 780  
 15 ACCGCCGGTG CCGCAGCTAA CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT 840  
 GAAACTCAAA GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC 900  
 20 GCAACGCGAA GAACCTTACC CAGGCTGGAC ATGCAGGTAG TAGAAGGGTG AAAGCCTAAC 960  
 GAGGTAGCAA TACCATCCTG CTCAGGTGCT GCATGGCTGT CGTCAGCTCG TGCCGTGAGG 1020  
 TGTGTTGGTTA AGTCCCGCAA CGAGCGCAAC CCCTGTCTTC AGTTACCAAC GGGTCATGCC 1080  
 25 GGGAACTCTG GAGAGACTGC CCAGGAGAAC GGGGAGGAAG GTGGGGATGA CGTCAAGTCA 1140  
 GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA AGCGCTGCAA 1200  
 30 ACCCGTAAGG GGGAGCCAAT CCCAAAAAAC CGGCCTCAGT TCAGATTGAG GTCTGCAACT 1260  
 CGACCTCATG AAGGCGGAAT CGCTAGTAAT CCCGGATCAG CACGCCGGGG TGAATACGTN 1320  
 CCCGGGCCTT GTACACACCG CCCGTCACAC CACGAAAGTT TGTGTACCT GAAGTCGTTG 1380  
 35 GCGCCAACCG CAAGGAGGCA GACGCCCACG GTATGACCGA TGATTGGG 1428

## (2) INFORMATION FOR SEQ ID NO: 2:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1407 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Nitrospira

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

5 TAATACATGC AAGTCGAGCG AGAAGACGTA GCAATACGTT TGTAAGCGG CGAACGGGTG 60  
 AGGAATACAT GGGTAGCCTA CCCTCGAGTG GGAATAACT AACCGAAAGG TTAGCTAATA 120  
 CCGCATACGG CTCCTGGTCT GCGGATCGGG AGAGAAAGCG ATACCGTGGG TATCGCGCTC 180  
 TTGGATGGGC TCATGTCCTA TCAGCTTGTT GGTGAGGTAA CGGCTCACCA AGGCTTCGAC 240  
 10 GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT 300  
 ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGACA GCCTGACGCA GCNACGCCGC 360  
 15 GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGC AGGGAAGATG GAACGGGTAA 420  
 CCGTTCGGAC GGTACCTGCA GAAGCAGCCA CGGCTAACTT CGTGCCAGCA GCCGCGGTAA 480  
 TACGAAGGTG GCAAGCGTTG TTCGGATTTA CTGGGCGTAC AGGGAGCGTA GCGGTTGGG 540  
 20 TAAGCCCTCC GTGAAATCTC CGGGCCTAAC CCGGAAAGTG CGGAGGGGAC TGCTCGGCTA 600  
 GAGGATGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG 660  
 GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATTTCTGA CGCTGAGGCT CGAAAGCGTG 720  
 25 GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCTT AAACGATGGA TACTAAGTGT 780  
 CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC 840  
 30 CGCAAGGTTG AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT 900  
 TAATTCGACG CAACGCGAAG AACCTTACCC AGGCTGGACA TGCAGGTAGT AGAAGGGTGA 960  
 AAGCCTAACG AGGTAGCAAT ACCATCCTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT 1020  
 35 GCCGTGAGGT GTTGGGTAA GTCCCGCAAC GAGCGCAACC CCTGTCTTCA GTTACCAACG 1080  
 GGTCATGCCG GGAACCTCTGG AGAGACTGCC CAGGAGAACG GGGGAGGAAG GTGGGGATGA 1140  
 40 CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA 1200  
 AGCGCTGCAA ACCCGTAAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG 1260  
 GTCTGCAACT CGACCTCATG AAGGCGGAAT CGCTAGTAAT CCCGGATCAG CACGCCGGGG 1320  
 45 TGAATACGTN CCCGGACCTT GTACACACCG CCCGTCACAC CACGAAAGTT TGTTGTACCT 1380  
 GAAGTCGTTG GCGCCAACCG CAAGGAG 1407

50 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 1500 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

15 TTGATCCTGG CTCAGAACGA ACCTGGCGG CGCGCCTAAT ACATGCAAGT CGAGCGAGAA 60  
 GACGTAGCAA TACGTTTGTA AAGCGGCGAA CGGGTGAGGA ATACATGGGT AACCTACCCT 120  
 CGAGTGGGGA ATAAGTAGCC GAAAGGTTAG CTAATACCGC ATACGACTCC TGGTCTGCGG 180  
 20 ATCGGGAGAG AAAGCGATAC CGTGGGTATC GCGCTCTTGG ATGGGCTCAT GTCCTATCAG 240  
 CTTGTTGGTG AGGTAACGGC TCACCAAGGC TTCGACGGGT AGCTGGTCTG AGAGGACGAT 300  
 CAGCCACACT GGCCTGCGA CACGGGCCAG ACTCCTACGG GAGGCAGCAG TAAGGAATAT 360  
 25 TGCGCAATGG GCGACAGCCT GACGCAGCNA CGCCGCGTGG GGGATGAAGG TCTTCGGATT 420  
 GTAAACCCCT TTCGGCAGGG AAGATGGAAC GGGTAACCGT TCGGACGGTA CCTGCAGAAG 480  
 30 CAGCCACGGC TAACCTCGTG CCAGCAGCCG CGGTAATACG AAGGTGGCAA GCGTTGTTTCG 540  
 GATTTACTGG GCGTACAGGG AGCGTAGGCG GTTGGGTAAG CCCTCCGTGA AATCTCCGGG 600  
 CCTAACCCGG AAAGTGCGGA GGGGACTGCT CGGCTAGAGG ATGGGAGAGG AGCGCGGAAT 660  
 35 TCCCGGTGTA GCGGTGAAAT GCGTAGAGAT CGGGAGGAAG GCCGGTGGCG AAGGCGGCGC 720  
 TCTGGAACAT TTCTGACGCT GAGGCTCGAA AGCGTGGGGA GCAAACAGGA TTAGATACCC 780  
 40 TGGTAGTCCA CGCCTTAAAC GATGGATACT AAGTGTCGGC GGGTTACCGC CGGTGCCGCA 840  
 GCTAACGCAT TAAGTATCCC GCCTGGGAAG TACGGCCGCA AGGTTGAAAC TCAAAGGAAT 900  
 TGACGGGGGC CCGCACAAGC GGTGGAGCAT GTGGTTTAAT TCGACGCAAC GCGAAGAACC 960  
 45 TTACCCAGGC TGGACATGCA GGTAGTAGAA GGGTGAAAGC CTAACGAGGT AGCAACACCA 1020  
 TCCTGCTCAG GTGCTGCATG GCTGTCGTCA GCTCGTGCCG TGAGGTGTTG GGTAAAGTCC 1080  
 50 CGCAACGAGC GCAACCCCTG TCTTCAGTTA CCAACGGGTC ATGCCGGGAA CTCTGGAGAG 1140  
 ACTGCCCAGG AGAACGGGGA GGAAGGTGGG GATGACGTCA AGTCAGCATG GCCTTTATGC 1200  
 CTGGGGCCAC ACACGTGCTA CAATGGCCGG TACAAAGCGC TGCAAACCCG TAAGGGGGAG 1260  
 55 CCAATCGCAA AAAACCGGCC TCAGTTCAGA TTGAGGTCTG CAACTCGACC TCATGAAGGC 1320

26

GGAATCGCTA GTAATCCCGG ATCAGCACGC CGGGGTGAAT ACGTNCCCGG GCCTTGTACA 1380  
 CACCGCCCGT CACACCACGA AAGTTTGTG TACCTGAAGT CGTTGGCGCC AACCGCAAGG 1440  
 5 GGGCAGACGC CCACGGTATG ACCGATGATT GGGGTGAAGT CGTAACAAGG TAACCGTAAC 1500

## (2) INFORMATION FOR SEQ ID NO: 4:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1420 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

30 CGAGAAGACG TAGCAATACG TTTGTAAAGC GCGAACGGG TGAGGAATAC ATGGGTAACC 60  
 TACCCCTCGAG TGGGGAATAA CTAACCGAAA GGTTAGCTAA TACCGCATAC GGCTCCTGGT 120  
 CTGCGGATCG GGAGAGAAAG CGATACCGTG GGTATCGCGC TCTTGATGG GCTCATGTCC 180  
 35 TATCAGCTTG TTGGTGAGGT AACGGCTCAC CAAGGCTTCG ACGGGTAGCT GGTCTGAGAG 240  
 GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC CTACGGGAGG CAGCAGTAAG 300  
 GAATATTGCG CAATGGGCGA CAGCCTGACG CAGCGACGCC GCGTTGGGGA TGAAAGTCTT 360  
 40 CCGATTGTAA ACCCCTTTCC GCAGGGAAGA TGGAACGGGT AACCCTTCGG ACGGTACCTG 420  
 CAGAAGCAGC CACGGCTAAC TTCGTGCCAG CAGCCGCGGT AATACGAAGG TGGCAAGCGT 480  
 45 TGTTTCGATT TACTGGGCGT ACAGGGAGCG TAGGCGGTTG GGTAAGCCCT CCGTGAAATC 540  
 TCCGGGCCTA ACCCGGAAAG TCGGAGGGG ACTGCTCGGC TAGAGGATGG GAGAGGAGCG 600  
 CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG AGGAAGGCCG GTGGCGAAGG 660  
 50 CGGCGCTCTG GAACATTTCT GACGCTGAGG CTCGAAAGCG TGGGGAGCAA ACAGGATTAG 720  
 ATACCCTGGT AGTCCACGCC TTAAACGATG GATACTAAGT GTCGGCGGGT TACCGCCGGT 780  
 55 GCCGCAGCTA ACGCATTAA GATCCCGCCT GGGAAGTACG GCCGCAAGGT TGAAACTCAA 840  
 AGGAATTGAC GGGGCCCCGC ACAAGCGGTG GAGCATGTGG TTTAATTCGA CGCAACGCGA 900

AGAACCTTAC CCAGGCAGGA CATGCAGGTA GTAGAAGGGT GAAAGCCTAA CGAGGTAGCA 960  
 ATACCATCCT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC GTGCCGTGAG GTGTTGGGTT 1020  
 5 AAGTCCCGCA ACGAGCGCAA CCCCTGTCTT CAGTTACCAA CGGGTCATGC CGGGAACCTCT 1080  
 GGAGAGACTG CCCAGGAGAA CGGGGAGGAA GGTGGGGATG ACGTCAAGTC AGCATGGCCT 1140  
 10 TTATGCCTGG GGCCACACAC GTGCTACAAT GGCCGGTACA AAGCGCTGCA AACCCGTAAG 1200  
 GGGGAGCCAA TCGCAAAAAA CCGGCCTCAG TTCAGATTGA GGTCTGCAAC TCGACCTCAT 1260  
 GAAGGCGGAA TCGCTAGTAA TCCCGGATCA GCACGCCGGG GTGAATACGT NCCGGGCCT 1320  
 15 TGTACACACC GCCCCTCACA CCACGAAAGT TTGTTGTACC TGAAGTCGTT GGCGCCAACC 1380  
 GCAAGGAGGC AGACGCCAC GGTATGACCG ATGATTGGGG 1420

20 (2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1505 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
 35 (A) ORGANISM: Nitrospira

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGAGTTTGAT CCTGGCTCAG AACGAACGCT GCGGGCGCGC CTAATACATG CAAGTCGAGC 60  
 GAGAAGACGT AGCAATACGT TTGTAAAGCG GCGAACGGGT GAGGAATACA TGGGTAATCT 120  
 45 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG CTTCTGAGTC 180  
 TTCGGGTTCG GAAGGAAAGC CGTACTGTGA GTGCGGCGCT CTTTGATGAG CTCATGTCCT 240  
 ATCAGCTTGT TGGTAGGGTA ACGGCCTACC AAGGCTTTGA CGGGTAGCTG GTCTGAGAGG 300  
 50 ACGATCAGCC AACTGGCAC TGCACACGG GCCAGACTCC TACGGGAGGC AGCAGTAAGG 360  
 AATATTGCGC AATGGGCGAA AGCCTGACGC AGCNACGCCG CGTGGGGGAT GAAGGTCTTC 420  
 55 GGATTGTAAA CCCCTTTCGG GAGGGAAGAT GGAGCGAGCA ATCGTTCGGA CGGTACCTCC 480  
 AGAAGCAGCC ACGGCCAACT TCGTGCCAGC AGCCGCGGTA ATACGAAGGT GGCAAGCGTT 540

GTTCGGATTC ACTGGGCGTA CAGGGTGTGT AGGCGGTTTG GTAAGCCTTC TGTTAAAGCT 600  
 TCGGGCCCAA CCCGAAAGC GCAGACGGTA CTGCCAGGCT AGAGGGTGGG AGAGGAGCGC 660  
 5 GGAATTCCCG GTGTAGCGGT GAAATGCGTA GAGATCGGGA GGAAGGCCGG TGGCGAAGGC 720  
 GGCGCTCTGG AACATACCTG ACGCTGAGAC ACGAAAGCGT GGGGAGCAAA CAGGATTAGA 780  
 10 TACCCTGGTA GTCCACGCCC TAAACTATGG ATACTAAGTG TCGGCGGGTT ACCGCCGGTG 840  
 CCGCAGCTAA CGCATTAAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA 900  
 GGAATTGACG GGGGCCCCGA CAAGCGGTGG AGCATGTGGT TTAATTCGAC GCAACGCGAA 960  
 15 GAACCTTACC CAGGTTGGAC ATGCACGTAG TAGAAAGGTG AAAGCCTGAC GAGGTAGCAA 1020  
 TACCAGCGTG CTCAGGTGCT GCATGGCTGT CGTCAGCTCG TGCCGTGAGG TGTGGGGTTA 1080  
 20 AGTCCCGCAA CGAGCGCAAC CCCTGCTTTC AGTTGCTACC GGGTCATGCC GAGCACTCTG 1140  
 AAAGGACTGC CCAGGATAAC GGGGAGGAAG GTGGGGATGA CGTCAAGTCA GCATGGCCTT 1200  
 TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA AGCGCTGCAA ACCCGTGAGG 1260  
 25 GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG GTCTGCAACT CGACCTCATG 1320  
 AAGGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG TGAATACGTN CCCGGGCCTT 1380  
 30 GTACACACCG CCCGTCACAC CACGAAAGCC TGTGTACCT GAAGTCGCCC AAGCCAACCG 1440  
 CAAGGAGGCA GGCGCCCACG GTATGGCCCG TGATTGGGGT GAAGTCGTAA CAAGGTAACC 1500  
 GTAAA 1505

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 1441 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

45

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

50

## (vi) ORIGINAL SOURCE:

## (A) ORGANISM: Nitrospira

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAGCGG CGAACGGGTG AGGAATACAT 60

5 GGGTAATCTA CCATCGAGTG GGGAATAACC AGCCGAAAGG TTGGCTAATA CCGCGTACGC 120  
 TTCCGAGTCT TCGGGCTTGG AAGGAAAGCC GCACTGTGAG TGCGGCGCTC TTTGATGAGC 180  
 TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC GGGTAGCTGG 240  
 TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT ACGGGAGGCA 300  
 10 GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCGACGCCGC GTGGGGGATG 360  
 AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCCAGCAA TCGTTCGGAC 420  
 GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA TACGAAGGTG 480  
 15 GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA NGCGGTTTGG TAAGCCTTCT 540  
 GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGAGGGTAC TGCCAGGCTA GAGGGTGGGA 600  
 20 GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG GAAGGCCGGT 660  
 GGCGAAGGCG GCGCTCTGGA ACATGCCTGA CGCTGAGACA CGAAAGCGTG GGGAGCAAAC 720  
 AGGATTAGAT ACCCTGGTAG TCCACGCCCT AACTATGGA TACTAAGTGT CGGCGGGTTA 780  
 25 CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC CGCAAGGTTG 840  
 AACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT TAATTGACG 900  
 30 CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA AAGNCTAACG 960  
 AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT GCCGTGAGGT 1020  
 GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG GGTGATGCCG 1080  
 35 AGCACTCTGA AAGGACTGCC CAGGATAACG GGGAGGAAGG TGGGGATGAC GTCAAGTCAG 1140  
 CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA GCGCTGCAAA 1200  
 40 CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG TCTGCAACTC 1260  
 GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT GAATACGTNC 1320  
 CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG AAGTCGCCCCA 1380  
 45 AGCCAACCGC AAGGAGGCAG GCGCCACGG TATGGCCGGT GATTGGGGTG AAGTCCTAAC 1440  
 A 1441

50 (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55

30

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

15	TAATACATGC AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAGCGG CGAACGGGTG	60
	AGGAATACAT GGGTAATCTA CCATCGAGTG GGAATAACC AACCGAAAGG TTGGCTAATA	120
	CCGCGTACGC TTCTGAGCCT TCGTGTCGG AAGGAAAGCC GTACTGTGAG TCGGCGCTC	180
20	TTTGATGAGC TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC	240
	GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT	300
	ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCNACGCCGC	360
25	GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCGAGCAA	420
	TCGTTCGGAC GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA	480
30	TACGAAGGTG GCAAGCGTTG CTTGGATTCA CTGGGCGTAC AGGGTGTGTA GCGGTTTGG	540
	TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGAAAAGCG CAGAGGGTAC TGCCAGGCTA	600
	GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG	660
35	GAAGGCCGGT GCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTGAGACA CGAAAACGTG	720
	GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAATATGGA TACTAAGTGT	780
40	CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAGGTACGGC	840
	CGCAAGGTTG AAATCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCTTGTGGTT	900
45	TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA	960
	AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT	1020
	GCCGTGAGGT GTTGGGTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG	1080
50	GGTCATGCCG AGCACTCTGA AAGGACTGCC CAGGATAACG GGGAGGAAGG TGGGGATGAC	1140
	GTCAAGTCAG CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA	1200
	GCGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAACC GGCCTCAGTT CAGATTGAGG	1260
55	TCTGCAACTC GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT	1320



31

GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG 1380

AAGTCGCCCCA AGCCAACCGC AAGGAGGCAG GCGCCCACGG TATGGC 1426

5 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1429 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

20 (A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

25 TAATACATGC AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAGCGG CGAACGGGTG 60

AGGAATACAT GGGTAATCTA CCATCGAGTG GGAATAACC AACCGAAAGG TTGGCTAATA 120

30 CCGCGTACGC CTCCGAGTCT TCGGGTTCGG AGGGAAGCT GCACTGTGAG TGTAGCGCTC 180

TTTGATGAGC TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC 240

GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT 300

35 ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCNACGCCGC 360

GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCGAGCAA 420

40 TCGTTCGGAC GTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA 480

TACGAAGGTG GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA GGCGGTTTGG 540

TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGGGGGTAC TGCCAGGCTA 600

45 GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG 660

GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTGAGACA CGAAAGCGTG 720

50 GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAGCTATGGA TACTAAGTGT 780

CGGCGGGTTA CCGCCGGTGC CGCAGCCAAC GCGTTAAGTA TCCCGCCTGG GAAGTACGGC 840

CGCAAGGTTG AACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT 900

55 TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA 960

32

AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT 1020  
 GCCGTGAGGT GTTGGGTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG 1080  
 5 GGTCATGCCG AGCACTCTGA AAGGACTGCC CAGGATAACG GGGGAGGAAG GTGGGGATGA 1140  
 CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA 1200  
 AACGCTGCAA ACCCGTGAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG 1260  
 10 GTCTGCAACT CGACCTCATG AAGGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG 1320  
 TGAATACGTN CCCGGGCCTT GTGCACACCG CCCGTCACAC CACGAAAGCC TGTGTACCT 1380  
 15 GAAGTCGCCC AAGCCAACCG CAAGGAGGCA GCGCCACG GTATGGCCG 1429

(2) INFORMATION FOR SEQ ID NO: 9:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1415 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

30 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Nitrospira

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGAGAAGGTG TAGCAATACA CTTGTAAAGC GGCGAACGGG TGAGGAATAC ATGGGTAATC 60  
 40 TACCATCGAG TGGGGAATAA CCAACCGAAA GGTGGCTAA TACCGCGTAC GCCTCCGAGT 120  
 CTCGGGTTC GGAGGGAAAG CTGCACTGTG AGTGTAGCGC TCTTTGATGA GCTCATGTCC 180  
 TATCAGCTTG TTGGTAGGGT AACGGCCTAC CAAGGCTTTG ACGGGTAGCT GGTCTGAGAG 240  
 45 GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC CTACGGGAGG CAGCAGTAAG 300  
 GAATATTGCG CAATGGGCGA AAGCCTGACG CAGCNACGCC GCGTGGGGGA TGAAGGTCTT 360  
 50 CGGATTGTAA ACCCCTTTTC GGAGGGAAGA TGGAGCGAGC AATCGTTCGG ACGGTACCTC 420  
 CAGAAGCAGC CACGGCCAAC TTCGTGCCAG CAGCCGCGGT AATACGAAGG TGGCAAGCGT 480  
 TGTTCGGATT CACTGGGCGT ACAGGGTGTG TAGGCGGTTT GGTAAGCCTT CTGTTAAAGC 540  
 55 TTCGGGCCCCA ACCCGGAAAG CGCAGAGGGT ACTGCCAGGC TAGAGGGTGG GAGAGGAGCG 600

33

CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG AGGAAGGCCG GTGGCGAAGG 660  
 CGGCGCTCTG GAACATACCT GACGCTGAGA CACGAAAGCG TGGGGAGCAA ACAGGATTAG 720  
 5 ATACCCTGGT AGTCCACGCC CTAAACTATG GATACTAAGT GTCGGCGGGT TACCGCCGGT 780  
 GCCGCAGCTA ACGCATTAAG TATCCCGCCT GGAAGTACG GCCGCAAGGT TGAAACTCAA 840  
 10 AGGAATTGAC GGGGGCCCCG ACAAGCGGTG GAGCATGTGG TTTAATTCGA CGCAACGCGA 900  
 AGAACCTTAC CCAGGTTGGA CATGCACGTA GTAGAAAGGT GAAAGCCTGA CGAGGTAGCA 960  
 ATACCAGCGT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC GTGCCGTGAG GTGTTGGGTT 1020  
 15 AAGTCCCGCA ACGAGCGCAA CCCCTGCTTT CAGTTGCTAC CGGGTCATGC CGAGCACTCT 1080  
 GAAAGGACTG CCCAGGATAA CGGGGAGGAA GGTGGGGATG ACGTCAAGTC AGCATGGCCT 1140  
 TTATGCCTGG GGCCACACAC GTGCTACAAT GGCCGGTATA AAACGCTGCA AATCCGTGAG 1200  
 20 GGGGAGCCAA TCGCAAAAAA CCGGCCTCAG TTCAGATTGA GGTCTGCAAC TCGACCTCAT 1260  
 GAAGGCGGAA TCGCTAGTAA TCGCGGATCA GCACGCCGCG GTGAATACGT NCCCGGGCCT 1320  
 25 TGTACACACC GCCCCTCACA CCACGAAAGC CTGTTGTACC TGAAGTCGCC CAAGCCAACC 1380  
 GCAAGGAGGC AGGCGCCAC GGTATGGCCG GTGAT 1415

## (2) INFORMATION FOR SEQ ID NO: 10:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1435 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

40

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Nitrospira

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

50 CCTAATACAT GCAAGTCGAT CGAGAAGGTG TAGCAATACA CTTGTAAAGC GGCGAACGGG 60  
 TGAGGAATAC ATGGGTAATC TACCATCGAG TGGGGAATAA CCAACCGAAA GGTTGGCTAA 120  
 TACCGCGTAC GCCTCCGAGT CTTCGGGTTC GGAGGGAAAG CTGCACTGTG AGTGTAGCGC 180  
 55 TCTTTGATGA GCTCATGTCC TATCAGCTTG TTGGTAGGGT AACGGCCTAC CAAGGCTTTG 240

34

ACGGGTAGCT GGTCTGAGAG GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC 300  
 CTACGGGAGG CAGCAGTAAG GAATATTGCG CAATGGGCGA AAGCCTGACG CAGCCACGCC 360  
 5 GCGTGGGGGA TGAAGGTCTT CGGATTGTAA ACCCCTTTTCG GGAGGGAAGA TGGAGCGAGC 420  
 AATCGTTCGG ACGGTACCTC CAGAAGCAGC CACGGCCAAC TTCGTGCCAG CAGCCGCGGT 480  
 10 AATACGAAGG TGGCAAGCGT TGTTCGGATT CACTGGGCGT ACAGGGTGTG TAGGCGGTTT 540  
 GGTAAGCCTT CTGTTAAAGC TTCGGGCCCCA ACCCGGAAAG CGCAGAGGGT ACTGCCAGGC 600  
 TAGAGGGTGG GAGAGGAGCG CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG 660  
 15 AGGAAGGCCG GTGGCGAAGG CGGCGCTCTG GAACATACCT GACGCTGAGA CACGAAAGCG 720  
 TGGGGAGCAA ACAGGATTAG ATACCCCTGGT AGTCCACGCC CTAAACTATG GATACTAAGT 780  
 GTCGGCGGGT TACCGCCGGT GCCGCAGCTA ACGCATTAAG TATCCCGCCT GGGAAAGTACG 840  
 20 GCCGCAAGGT TGAAACTCAA AGGAATTGAC GGGGGCCCCG ACAAGCGGTG GAGCATGTGG 900  
 TTTAATTCGA CGCAACGCGA AGAACCTTAC CCAGGTTGGA CATGCACGTA GTAGAAAGGT 960  
 25 GAAAGCCTGA CGAGGTAGCA ATACCAGCGT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC 1020  
 GTGCCGTGAG GTGTTGGGTT AAGTCCCGCA ACGAGCGCAA CCCCTGCTTT CAGTTGCTAC 1080  
 CGGGTCATGC CGAGCACTCT GAAAGGACTG CCCAGGATAA CGGGGAAGGA AGGTGGGGAT 1140  
 30 GACGTCAAGT CAGCATGGCC TTTATGCCTG GGGCCACACA CGTGCTACAA TGGCCGGTAC 1200  
 AAAACGCTGC AAACCCGTGA GGGGGAGCCA ATCGCAAAAA ACCGGCCTCA GTTCAGATTG 1260  
 35 AGGTCTGCAA CTCGACCTCA TGAAGGCGGA ATCGTAGTA ATCGCGGATC AGCAGCGCGC 1320  
 GGTGAATACG TNCCCGGGCC TTGTACACAC CGCCCGTCAC ACCACGAAAG CCTGTTGTAC 1380  
 40 CTGAAGTCGC CCAAGCCAAC CGCAAGAAGG CAGGCGCCCA CGGTATGGCC GGTGA 1435

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 1437 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

50

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

5	AATACATGCA AGTCGATCGA GAAGGTGTAG CAATACACTT GTAAAGCGGC GAACGGGTGA	60
	GGAATACATG GGTAATCTAC CATCGAGTGG GGAATAACCA ACCGAAAGGT TGGCTAATAC	120
	CGCGTACGCC TCCGAGTCTT CGGGTTCGGA GGGAAAGCTG CACTGTGAGT GTAGCGCTCT	180
10	TTGATGAGCT CATGTCCTAT CAGCTTGTTG GTAGGGTAAC GGCCTACCAA GGCTTTGACG	240
	GGTAGCTGGT CTGAGAGGAC GATCAGCCAC ACTGGCACTG CGACACGGGC CAGACTCCTA	300
15	CGGGAGGCAG CAGTAAGGAA TATTGCGCAA TGGGCGAAAG CCTGACGCAG CCACGCCGCG	360
	TGGGGGATGA AGGTCTTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG AGCGAGCAAT	420
	CGTTCGGACG GTACCTCCAG AAGCAGCCAC GGCCAACTTC GTGCCAGCAG CCGCGGTAAT	480
20	ACGAAGGTGG CAAGCGTTGT TCGGATTCAC TGGGCGTACA GGGTGTGTAG GCGGTTTGGT	540
	AAGCCTTCTG TTAAAGCTTC GGGCCCAACC CGGAAAGCGC AGAGGGTACT GCCAGGCTAG	600
25	AGGGTGGGAG AGGAGCGCGG AATTCCCGGT GTAGCGGTGA AATGCGTAGA GATCGGGAGG	660
	AAGGCCGGTG GCGAAGGCGG CGCTCTGGAA CATACTGAC GCTGAGACAC GAAAGCGTGG	720
	GGAGCAAACA GGATTAGATA CCCTGGTAGT CCACGCCCTA AACTATGGAT ACTAAGTGTC	780
30	GGCGGGTTAC CGCCGGTGCC GCAGCTAACG CATTAAGTAT CCCGCCTGGG AAGTACGGCC	840
	GCAAGGTTGA AACTCAAAGG AATTGACGGG GGCCCGCACA AGCGGTGGAG CATGTGGTTT	900
35	AATTCGACGC AACGCGAAGA ACCTTACCCA GGTGGACAT GCACGTAGTA NAAAGGTGAA	960
	AGCCTGACGA GGTAGCAATA CCAGCGTGCT CAGGTGCTGC ATGGCTGTCT TCAGCTCGTG	1020
	CCGTGAGGTG TTGGGTAAAG TCCCGCAACG AGCGCAACCC CTGCTTTCAG TTGCTACCGG	1080
40	GTCATGCCGA ACACTCTGAA AGGACTGCCC AGGATAACGG GGAAGGAAGG TGGGGATGAC	1140
	GTCAAGTCAG CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA	1200
45	GCGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG	1260
	TCTGCAACTC GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT	1320
	GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTAACCTG	1380
50	AAGTCGCCCA AGCCAACCGC AAGGAGGCAG GCGCCACGG TATGGCCGGT GATGGGG	1437

## (2) INFORMATION FOR SEQ ID NO: 12:

- 55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1437 base pairs  
 (B) TYPE: nucleic acid

36

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10 (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospira

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

	AATACATGCA AGTCGATCGA NAAGGTGTAG CAATACACTT GTAAAGCGGC GAACGGGTGA	60
20	GGAATACATG GGTAATCTAC CATCGAGTGG GGAATAACCA ACCGAAAGGT TGGCTAATAC	120
	CGCGTACGCC TCCGAGTCTT CGGGTTCGGA GGGAAAGCTG CACTGTGAGT GTAGCGCTCT	180
	TTGATGAGCT CATGTCCTAT CAGCTTGTG GTAGGGTAAC GGCCTACCAA GGCTTTGACG	240
25	GGTATCTGGT CTGAGAGGAC GATCAGCCAC ACTGGCACTG CGACACGGGC CAGACTCCTA	300
	CGGGAGGCAG CAGTAAGGAA TATTGCGCAA TGGGCGAAAC CCNGACGCAG CCACGCCGCG	360
30	TGGGGGATGA AGGTCTTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG AACGAGCAAT	420
	CGTTCGGACG GTACCTCCAG AAGCAGCCAC GGCCAACTTC GTGCCAGCAG CCGCGGTAAT	480
	ACGAAGGTGG CAAGCGTTGT TCGGATTCAC TGGGCGTACA GGGTGTGTAG GCGGTTTGGT	540
35	AAGCCTTCTG TTAAAGCTTC GGGCCCAACC CGGAAAGCGC AGAGGGTACT GCCAGGCTAG	600
	AGGGTGGGAG AGGAGCGCGG AATTCCCGGT GTAGCGGTGA AATGCGTAGA GATCGGGAGG	660
40	AAGGCCGGTG GCGAAGGCGG CGCTCTGGAA CATACTGAC GCTGAGACAC GAAAGCGTGG	720
	GGNGCAAACA GGATTAGATA CCCTGGTAGT CCACGCCCTA AACTATGGAT ACTAAGTGTC	780
	GGCGGGTTAC CGCCGGTGCC GCAGCTAACG CATTAAGTAT CCCGCCTGGG AAGTACGGCC	840
45	GCAAGGTTGA AACTCAAAGG GATTGACGGG GGCCCGCACA AGCGGTGGGG CATGTGGTTT	900
	AATTCGACGC AACGCGAAGA ACCTTACCCA GGTTGGACAT GCACGTAGTN GAAAGGTGAA	960
50	AGCCTGACGA GGTAGCAATA CCAGCGTGCT CAGGTGCTGC ATGGCTGTCTG TCAGCTCGTG	1020
	CCGTGAGGTG TTGGGTTAAG TCCCGCAACG AGCGCAACCC CTGCTTTCAG TTGCTACCGG	1080
	GTCATGCCGA AACTCTGAA AGGACTGCCC AGGATAACGG GGAAGGAAGG TGGGGATGAC	1140
55	GTCAAGTCAG CATGGCCTTT ATACCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA	1200
	ACGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAACC GGCCTCAGTT CAGATTGAGG	1260

TCTGCAACTC GACCTCATGA ATGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT 1320  
 GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG 1380  
 5 AAGTCGCCCCA AGCCAACCGC AAGGAGGCAG GCGCCACGCG TATGGCCGGT GATGGGG 1437

## (2) INFORMATION FOR SEQ ID NO: 13:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1435 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Nitrospira

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATACATGC AAGTCGATCG ANAAGGTGTA GCAATACACT TGTAAGCGG CGAACGGGTG 60  
 30 AGGAATACAT GGGTAATCTA CCATCGAGTG GGAATAACC AACCGAAAGG TTGGCTAATA 120  
 CCGCGTACGC TTCCGAGTCT TCGGGCTTGG AAGGAAAGCC GCACTGTGAG TGCGGCGCTC 180  
 35 TTTGATGAGC TCATATCCTA TCANCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC 240  
 GGGTATCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT 300  
 ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA CCCNGACGCA GCCACGCCGC 360  
 40 GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAACGAGCAA 420  
 TCGTTCGGAC GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA 480  
 45 TACGAAGGTG GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA GGCGGTTTGG 540  
 TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGAGGGTAC TGCCAGGCTA 600  
 GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG 660  
 50 GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTCAGACA CGAAAGCGTG 720  
 GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AACTATGGA TACTAAGTGT 780  
 55 CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC 840  
 CGCAAGGTTG AAACCTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT 900

TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA 960  
 AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT 1020  
 5 GCCGTGAGGT GTTGGGTTAA GTCCC GCAAC GAGCGCAACC CCTGCTTTCA GTT'GCTGCCG 1080  
 GGTCATGCCG AACACTCTGA AAGGACTGCC CAGGATAACG GGAAGGAAG GTGGGGATGA 1140  
 10 CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA 1200  
 AACGCTGCAA ACCCGTGAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCANATTGAG 1260  
 GTCTGCAACT CGACCTCATG AATGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG 1320  
 15 TGAATACGTN CCCGGGCCTT GTACACGCCG CCCGTCACAC CACGAAAGCC TGTGTACCT 1380  
 GAAGTCGCCC AAGCCAACCG CAAGGAGGCA NGCGCCCACG GTATGGCCGG TGATG 1435

20 (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 18 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "oligonucleotide primer"

30 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

40 CGGGAGGGAA GATGGAGC

18

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:  
 45 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "Oligonucleotide primer"

(iii) HYPOTHETICAL: NO

55 (iv) ANTI-SENSE: NO



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

5 CCAACCCGGA AAGCGCAGAG

20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide primer"

(iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGCCTGGCAG TACCCTCT

18

(2) INFORMATION FOR SEQ ID NO: 17:

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

40 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

45 (A) ORGANISM: Nitrococcus mobilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

50 CAGCCGGGAG GAAAAGCA

18

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

10 (A) ORGANISM: Magnetobacterium bavaricum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

15

TG TAGGGAAA GATGATGA

18

(2) INFORMATION FOR SEQ ID NO: 19:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

30

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrobacter hamburgensis

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

40 TGTGCGGGAA GATAATGA

18

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

45

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

50

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospina gracilis

41

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 CGGGTGGGAA GAACAAAA

18

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira marina

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CATGAGGAAA GATAAAGT

18

30 (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

45 (A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

50

CGGCAGGGAA GATGGAAC

18

(2) INFORMATION FOR SEQ ID NO: 23:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid

42

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10 (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospira

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CGGGAGGGAA GATGGAGC

18

20 (2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
35 (A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

40 CCGCAGGGAA GATGGAAC

18

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:  
45 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55 (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

5

CGGGAGGGAA GATGGAAC

18

(2) INFORMATION FOR SEQ ID NO: 26:

10

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

20

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrobacter

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CGTGCGGGAA GATAATGA

18

30

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA (genomic)

40

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

45

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CGGCAGGGAA GATGGAAC

18

(2) INFORMATION FOR SEQ ID NO: 28:

55

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs

44

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Nitrospira moscoviensis*

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CGGGAGGGAA GATGGACG

18

20 (2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

25

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

35

(A) ORGANISM: *Nitrococcus mobilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

40

TCAACCTGGG AATTGCATCC

20

(2) INFORMATION FOR SEQ ID NO: 30:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

55

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

45

(A) ORGANISM: Magnetobacterium bavaricum

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TCAACCCGGG AATTGCCTTG

20

10 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

20

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrobacter hamburgensis

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

30 TCAACTCCAG AACTGCCTTT

20

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35

40 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

45 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospina gracilis

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

TCAACCGTGG AATTGCGTTT

20

55 (2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

46

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

10

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospina marina

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTAACCGGGA AAGGTCGAGA

20

20

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA (genomic)

30

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

35

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CTAACCCGGA AAGTGCGGAG

20

(2) INFORMATION FOR SEQ ID NO: 35:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

55

(iv) ANTI-SENSE: NO



47

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospira

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCAACCCGAA AAGCGCAGAG

20

10 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

20 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospira

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

30

CCAACCCGGA AAGCGCAGAG

20

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35

40

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

45 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrobacter

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TCAACTCCAG AACTGCCTTT

20

55

(2) INFORMATION FOR SEQ ID NO: 38:

48

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
5       (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 10       (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:  
            (A) ORGANISM: Nitrospira moscoviensis
- 15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

20   CCAACCCGGA AAGCGCAGAG

20

(2) INFORMATION FOR SEQ ID NO: 39:

- 25       (i) SEQUENCE CHARACTERISTICS:  
            (A) LENGTH: 18 base pairs  
            (B) TYPE: nucleic acid  
            (C) STRANDEDNESS: single  
            (D) TOPOLOGY: linear
- 30       (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- 35       (vi) ORIGINAL SOURCE:  
            (A) ORGANISM: Nitrococcus mobilis

40       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGCCAAACAG TATCGGAT

18

45   (2) INFORMATION FOR SEQ ID NO: 40:

- 50       (i) SEQUENCE CHARACTERISTICS:  
            (A) LENGTH: 18 base pairs  
            (B) TYPE: nucleic acid  
            (C) STRANDEDNESS: single  
            (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 55       (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Magnetobacterium bavaricum

5

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AGTTAAACAG TTTTCAAG

18

10

- (2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15

- (ii) MOLECULE TYPE: DNA (genomic)

20

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

25

- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrobacter hamburgensis

30

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

AGACCTTCAG TATCAAAG

18

35

- (2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

40

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

45

- (iv) ANTI-SENSE: NO

- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Nitrospina gracilis

50

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

55

AGCCGAATAG TTTCAAAC

18

- (2) INFORMATION FOR SEQ ID NO: 43:

50

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 18 base pairs  
    (B) TYPE: nucleic acid  
5      (C) STRANDEDNESS: single  
        (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 10    (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:  
15        (A) ORGANISM: Nitrospina marina

20    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:  
AGCTGAATAG TTCCTCTC

18

(2) INFORMATION FOR SEQ ID NO: 44:

- 25    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 18 base pairs  
        (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single  
30        (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- 35        (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:  
            (A) ORGANISM: Nitrospira

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

45    AGCCGAGCAG TCCCCTCC

18

(2) INFORMATION FOR SEQ ID NO: 45:

- 50    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 18 base pairs  
        (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single  
        (D) TOPOLOGY: linear
- 55        (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

51

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

10 AGCCTGGCAG TACCCTCT

18

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

20 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

AGCCTGGCAG TACCCCCT

18

35 (2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

45

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

55

AGCCTGGCAG TACCGTCT

18 .

52

## (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Nitrobacter

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AGATCCTCAG TATCAAAG

18

## (2) INFORMATION FOR SEQ ID NO: 49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Nitrospira moscoviensis

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGCCTGGCAG TACCCTCT

18

## (2) INFORMATION FOR SEQ ID NO: 50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Oligonucleotide primer"

53

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

10 CCTGTGCTCC ATGCTCCG

18

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrobacter hamburgensis

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CCTGTGCTCC ATGCTCCG

18

35 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

45 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

50 (A) ORGANISM: Nitrospina gracilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

55

CCTGTGCAAG GGCCCCGA

18

54

## (2) INFORMATION FOR SEQ ID NO: 53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Nitrococcus mobilis*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CCTGTCATCC GGTTC

18

## (2) INFORMATION FOR SEQ ID NO: 54:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Nitrospira moscoviensis*

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CCTGAGCACG CTGGTATT

18

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO



55

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

5 (A) ORGANISM: Nitrospina marina

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10

CCTGAGCTCG CTCCCCTT

18

(2) INFORMATION FOR SEQ ID NO: 56:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

25

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Magnetobacterium bavaricum

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

CCTGTGCAAG CTCTCCCT

18

35

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

40

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

45

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

50

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CCTGAGCAGG ATGGTATT

18

56

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nitrospira

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CCTGAGCACG CTGGTATT

18

25 (2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

40 (A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

45

CCTGAGCAGG ATGGTGTT

18

## CLAIMS

1. A consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum.
2. An oligonucleotide primer for PCR amplification of *Nitrospira* DNA, said primer comprising  
5 at least 12 nucleotides having a sequence selected from the group consisting of:
  - (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and
  - (ii) a DNA sequence having at least 92 % identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.
3. The oligonucleotide primer of claim 2, wherein said primer has a length of 12 to 50  
10 nucleotides.
4. The oligonucleotide primer of claim 2, wherein said primer has a length of 12 to 22 nucleotides.
5. The oligonucleotide primer of claim 2, wherein said primer sequence is selected from the group consisting of SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16.
- 15 6. A primer pair for PCR amplification of *Nitrospira* DNA, said primer pair comprising:
  - (a) a first oligonucleotide of at least 12 nucleotides having a sequence selected from one strand of a bacterial 16S rDNA gene; and
  - (b) a second oligonucleotide of at least 12 nucleotides having a sequence selected from the other strand of said 16S rDNA gene downstream of said first oligonucleotide sequence; wherein at  
20 least one of said first and second oligonucleotides is selected from the group consisting of:
    - (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and
    - (ii) a DNA sequence having at least 92 % identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.
7. The primer pair of claim 6, wherein said first and second oligonucleotide primers  
25 independently have lengths of 12 to 50 nucleotides.
8. The primer pair of claim 6, wherein said first and second oligonucleotide primers independently have lengths of 12 to 22 nucleotides.
9. The primer pair of claim 6, wherein said first oligonucleotide primer sequence is selected from the group consisting of SEQ ID NO: 14 and SEQ ID NO: 15, and said second oligonucleotide  
30 primer sequence is SEQ ID NO: 16.
10. A probe for detecting *Nitrospira* DNA, said probe comprising at least 12 nucleotides having a sequence selected from the group consisting of:
  - (i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and
  - (ii) a DNA sequence having at least 92 % identity with any one of SEQ ID NO: 1 to SEQ  
35 ID NO: 13.

11. The probe of claim 10, wherein said probe has a length of 15 to 50 nucleotides.
12. The probe of claim 10, wherein said probe has a length of 15 to 22 nucleotides.
13. A kit comprising:
  - at least one primer according to claim 2;
  - 5 at least one primer pair according to claim 6; or
  - at least one probe according to claim 10.
14. The kit of claim 13, wherein said kit further includes reagents selected from the group consisting of buffers, salts, detergents, nucleotides and thermostable polymerase.
15. A method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:
  - 10 (a) lysing cells in said sample to release genomic DNA;
  - (b) contacting denatured genomic DNA from step (a) with a primer pair according to claim 6;
  - (c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and
  - 15 (d) detecting said amplification product.
16. The method according to claim 15, wherein said amplification product has a length of 50 to 1,400 bps.
17. A method of quantitating the level of a *Nitrospira* species in a sample, said method comprising the steps of:
  - 20 (a) lysing cells in said sample to release genomic DNA;
  - (b) contacting denatured genomic DNA from step (a) with a primer pair according to claim 6;
  - (c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and
  - 25 (d) detecting said amplification product and quantitating the level of said product by comparison with at least one reference standard.
18. The method according to claim 17, wherein said amplification product has a length of 50 to 1,400 bps.
19. A method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:
  - 30 (a) lysing cells in said sample to release genomic DNA;
  - (b) contacting denatured genomic DNA from step (a) with a labelled probe according to claim 4 under conditions which allow hybridisation of said genomic DNA said probe;
  - (c) separating hybridised labeled probe and genomic DNA from unhybridised labeled probe; and
  - 35 (d) detecting said labeled probe-genomic DNA hybrid.

20. A method of detecting cells of a *Nitrospira* species in a sample, said method comprising the steps of:

- (a) treating cells in said sample to fix cellular contents;
- (b) contacting said fixed cells from step (a) with a labeled probe according to claim 10
- 5 under conditions which allow said probe to hybridise with RNA within said fixed cell;
- (c) removing unhybridised probe from said fixed cells; and
- (d) detecting said labeled probe-RNA hybrid.

Ridout & Maybee  
Suite 2400  
One Queen Street East  
Toronto, Canada M5C 3B1  
Patent Agents of the Applicant

1/17

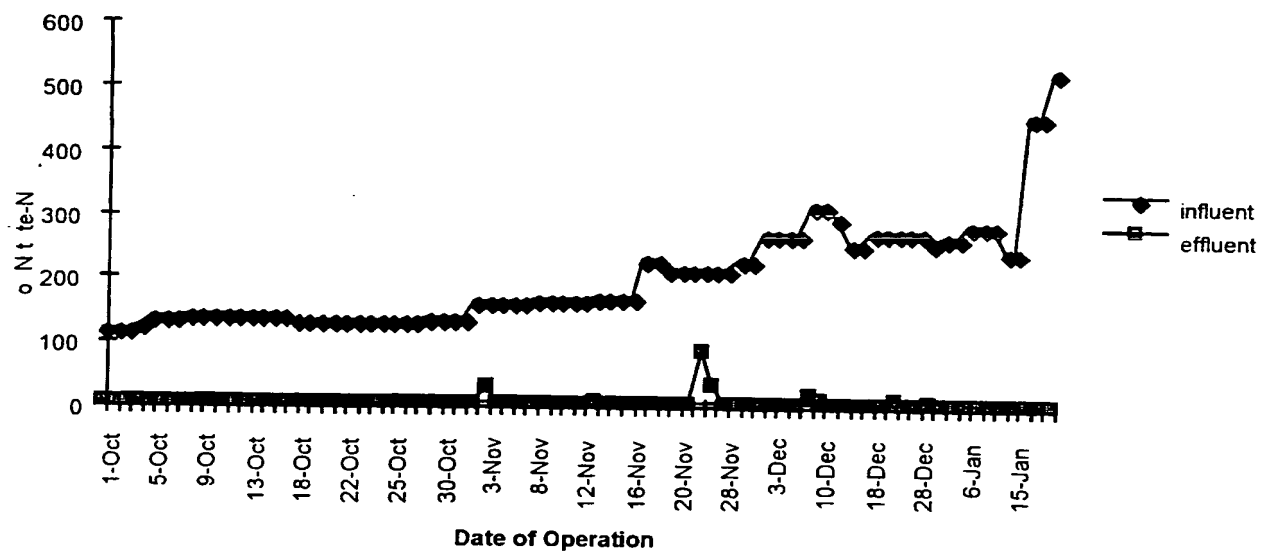


Fig. 1

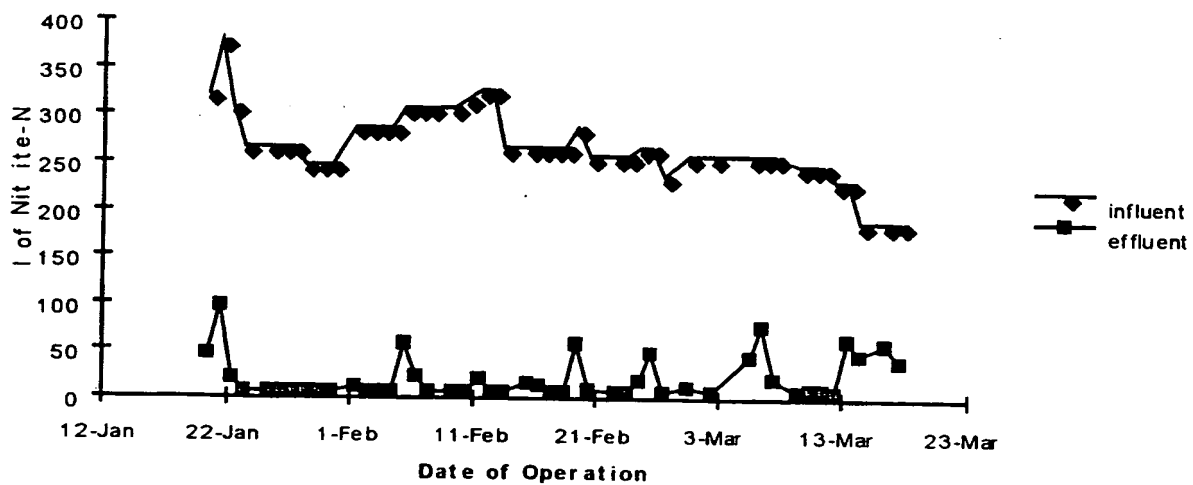


Fig. 2

2/17

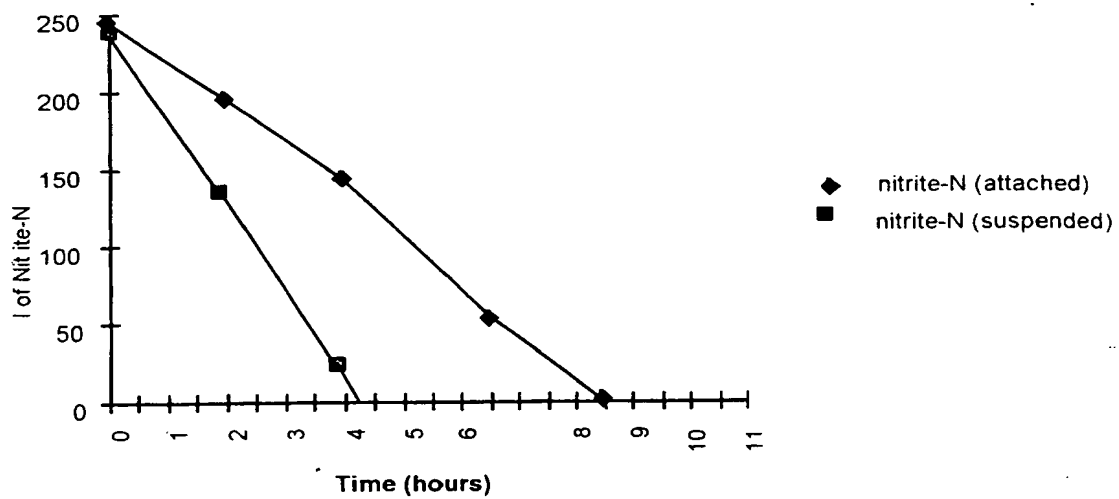


Fig. 3

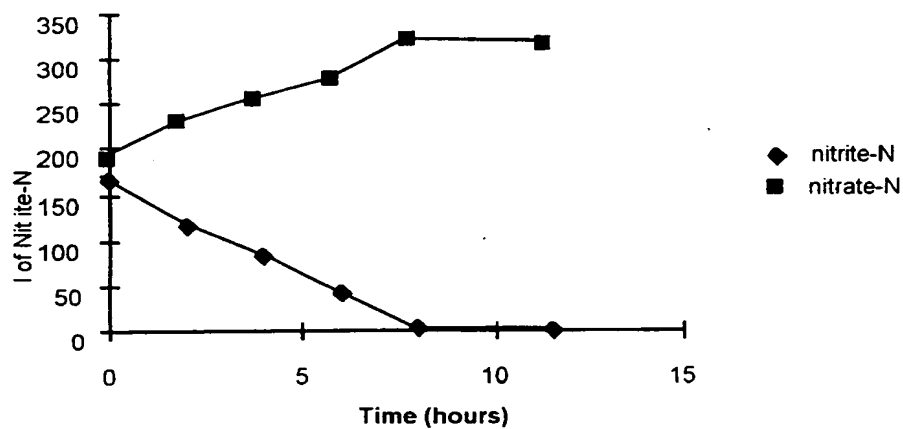
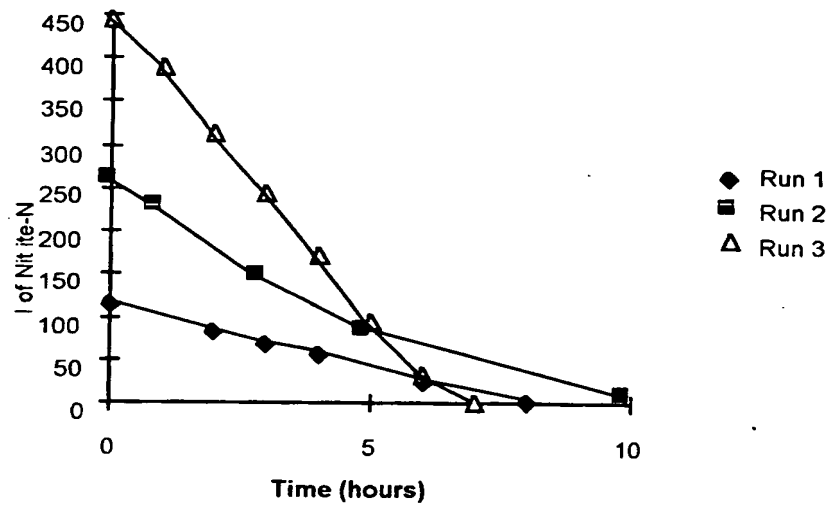
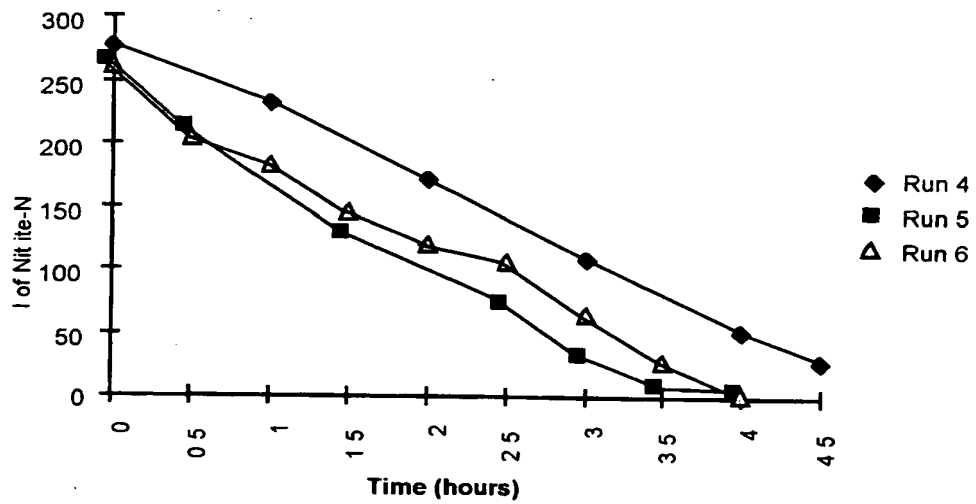


Fig. 4

3/17

*Fig. 5**Fig. 6*



4/17

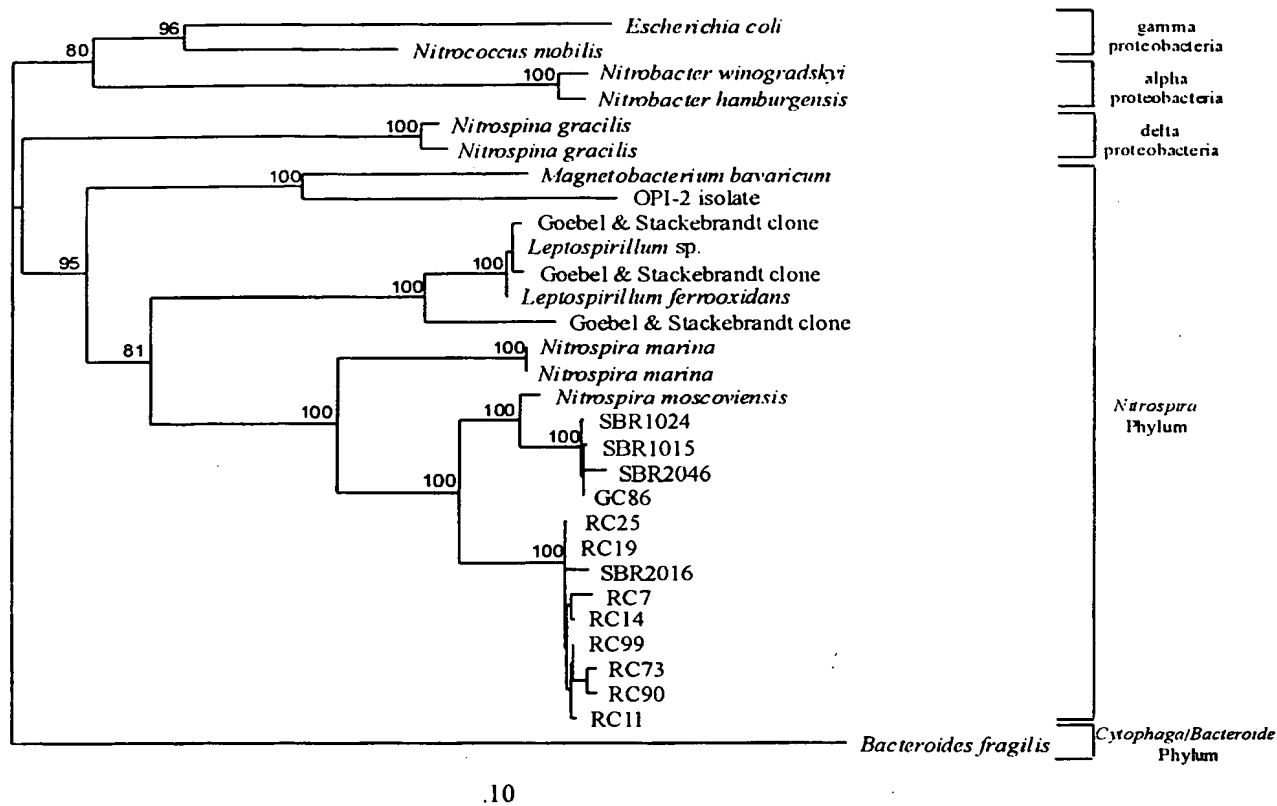


Fig. 7

5/17

```

[      1                                          50 ]
SBR1024-----
SBR1015-----
GC86  ----- --TCGACCTG CAGGCGGCCG CACTAGTGAT
SBR2046-----
RC25  -----GC TCTCCCATAT GGTCGACCTG CAGGCGGCCG CACTAGTGAT
RC19  -----
SBR2016-----
RC7   -----
RC14  -----
RC99  -----
RC11  -----
RC73  -----
RC90  -----

[      51                                          100 ]
SBR1024-----
SBR1015----- --TAATACAT
GC86  TAGAGTTTGA TCCTGGCTCA GAACGAACGC TGGCGGCGCG CCTAATACAT
SBR2046-----
RC25  TAGAGTTTGA TCCTGGCTCA GAACGAACGC TGGCGGCGCG CCTAATACAT
RC19  -----
SBR2016----- --TAATACAT
RC7   ----- --TAATACAT
RC14  -----
RC99  ----- CCTAATACAT
RC11  ----- --AATACAT
RC73  ----- --AATACAT
RC90  ----- --TAATACAT

[      101                                          150 ]
SBR1024-CAAGTCGAG CGAGAAGACG TA.....GCAA...TA
SBR1015GCAAGTCGAG CGAGAAGACG TA.....GCAA...TA
GC86  GCAAGTCGAG CGAGAAGACG TA.....GCAA...TA
SBR2046----- CGAGAAGACG TA.....GCAA...TA
RC25  GCAAGTCGAG CGAGAAGACG TA.....GCAA...TA
RC19  --AAGTCGAG CGAGAAGGTG TA.....GCAA...TA
SBR2016GCAAGTCGAG CGAGAAGGTG TA.....GCAA...TA
RC7   GCAAGTCGAG CGAGAAGGTG TA.....GCAA...TA
RC14  ----- CGAGAAGGTG TA.....GCAA...TA
RC99  GCAAGTCGAT CGAGAAGGTG TA.....GCAA...TA
RC11  GCAAGTCGAT CGAGAAGGTG TA.....GCAA...TA
RC73  GCAAGTCGAT CGANAAGGTG TA.....GCAA...TA
RC90  GCAAGTCGAT CGANAAGGTG TA.....GCAA...TA

```

Fig. 8

6/17

```

[ 151                                     200 ]
SBR1024CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
SBR1015CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAGCCT
GC86 CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
SBR2046CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
RC25 CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC19 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
SBR2016CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC7 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC14 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC99 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC11 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC73 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
RC90 CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT

[ 201                                     250 ]
SBR1024ACCTTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG
SBR1015ACCCTCGAGT GGGGAATAAC TAACCGAAAG GTTAGCTAAT ACCGCATACG
GC86 ACCCTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG
SBR2046ACCCTCGAGT GGGGAATAAC TAACCGAAAG GTTAGCTAAT ACCGCATACG
RC25 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC19 ACCATCGAGT GGGGAATAAC CAGCCGAAAG GTTGGCTAAT ACCGCGTACG
SBR2016ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC7 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC14 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC99 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC11 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC73 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
RC90 ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG

[ 251                                     300 ]
SBR1024ACTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ....GTG.
SBR1015GCTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ....GTG.
GC86 ACTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ....GTG.
SBR2046GCTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ....GTG.
RC25 CTTCTGAGTC .TTC..GGGT TCGGAAGGAA AGCCGTACT. ....GTG.
RC19 CTTCCGAGTC .TTC..GGGC TTGGAAGGAA AGCCGCACT. ....GTG.
SBR2016CTTCTGAGCC .TTC..GTGT TCGGAAGGAA AGCCGTACT. ....GTG.
RC7 CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ....GTG.
RC14 CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ....GTG.
RC99 CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ....GTG.
RC11 CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ....GTG.
RC73 CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ....GTG.
RC90 CTTCCGAGTC .TTC..GGGC TTGGAAGGAA AGCCGCACT. ....GTG.

```

*Fig. 8 (continued)*

7/17

```

[ 301 350 ]
SBR1024.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
SBR1015.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
GC86 .....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
SBR2046.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
RC25 .....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC19 .....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
SBR2016.....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC7 .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC14 .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC99 .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC11 .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC73 .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC90 .....AGTGC GGCGCTCTTT GATGAGCTCA TATCCTATCA NCTTGTTGGT

[ 351 400 ]
SBR1024GAGGTAACGG CTCACCAAGG CTTGACGGG TAGCTGGTCT GAGAGGACGA
SBR1015GAGGTAACGG CTCACCAAGG CTTGACGGG TAGCTGGTCT GAGAGGACGA
GC86 GAGGTAACGG CTCACCAAGG CTTGACGGG TAGCTGGTCT GAGAGGACGA
SBR2046GAGGTAACGG CTCACCAAGG CTTGACGGG TAGCTGGTCT GAGAGGACGA
RC25 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC19 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
SBR2016AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC7 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC14 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC99 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC11 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC73 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC90 AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA

[ 401 450 ]
SBR1024TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR1015TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
GC86 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR2046TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC25 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC19 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR2016TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC7 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC14 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC99 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC11 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC73 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC90 TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA

```

Fig. 8 (continued)

8/17

```

[ 451 500 ]
SBR1024GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
SBR1015GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
GC86 GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
SBR2046GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC GACGCCGCGT
RC25 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
RC19 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC GACGCCGCGT
SBR2016GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
RC7 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
RC14 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
RC99 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC CACGCCGCGT
RC11 GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC CACGCCGCGT
RC73 GTAAGGAATA TTGCGCAATG GGC.GAAACC CNGACGCAGC CACGCCGCGT
RC90 GTAAGGAATA TTGCGCAATG GGC.GAAACC CNGACGCAGC CACGCCGCGT

[ 501 550 ]
SBR1024GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAAGATGG
SBR1015GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAAGATGG
GC86 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAAGATGG
SBR2046TGGGGATGAA AGTC.TTCCG ATTGTAAACC CCTTTCGGCA GGGAAAGATGG
RC25 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC19 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
SBR2016GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC7 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC14 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC99 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC11 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC73 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG
RC90 GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAAGATGG

[ 551 600 ]
SBR1024AACGG.....GTAA.....CCGTTCG GACGGTACCT GCAGAAGCAG
SBR1015AACGG.....GTAA.....CCGTTCG GACGGTACCT GCAGAAGCAG
GC86 AACGG.....GTAA.....CCGTTCG GACGGTACCT GCAGAAGCAG
SBR2046AACGG.....GTAA.....CCGTTCG GACGGTACCT GCAGAAGCAG
RC25 AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC19 AGCCA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
SBR2016AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC7 AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC14 AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC99 AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC11 AGCGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC73 AACGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG
RC90 AACGA.....GCAA.....TCGTTCG GACGGTACCT CCAGAAGCAG

```

*Fig. 8 (continued)*

9/17

```

[ 601                                     650 ]
SBR1024CCACGGCTAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
SBR1015CCACGGCTAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
GC86 CCACGGCTAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
SBR2046CCACGGCTAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC25 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC19 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
SBR2016CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC7 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC14 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC99 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC11 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC73 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG
RC90 CCACGGCCAA CTTTCGTGCCA GCAGCCGCGG TAATACGAAG GTGGCAAGCG

[ 651                                     700 ]
SBR1024TTGTTTCGGAT TTACTGGGCG TACAGGGAGC GTAGGCGGTT GGGTAAGCCC
SBR1015TTGTTTCGGAT TTACTGGGCG TACAGGGAGC GTAGGCGGTT GGGTAAGCCC
GC86 TTGTTTCGGAT TTACTGGGCG TACAGGGAGC GTAGGCGGTT GGGTAAGCCC
SBR2046TTGTTTCGGAT TTACTGGGCG TACAGGGAGC GTAGGCGGTT GGGTAAGCCC
RC25 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC19 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
SBR2016TTGCTTGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC7 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC14 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC99 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC11 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC73 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT
RC90 TTGTTTCGGAT TCACTGGGCG TACAGGGTGT GTAGGCGGTT TGGTAAGCCT

[ 701                                     750 ]
SBR1024TCCGTGAAAT CTCCGGGCCT AACCCGGAAG GTGCGGAGGG GACTGCTCGG
SBR1015TCCGTGAAAT CTCCGGGCCT AACCCGGAAG GTGCGGAGGG GACTGCTCGG
GC86 TCCGTGAAAT CTCCGGGCCT AACCCGGAAG GTGCGGAGGG GACTGCTCGG
SBR2046TCCGTGAAAT CTCCGGGCCT AACCCGGAAG GTGCGGAGGG GACTGCTCGG
RC25 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC19 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
SBR2016TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC7 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC14 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC99 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC11 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC73 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG
RC90 TCTGTTAAAG CTTTCGGGCC AACCCGGAAG GCGCAGAGGG TACTGCCAGG

```

*Fig. 8 (continued)*

10/17

[	751				800 ]
SBR1024	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR1015	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
GC86	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR2046	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC25	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC19	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR2016	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC7	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC14	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC99	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC11	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC73	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC90	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
[	801				850 ]
SBR1024	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
SBR1015	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
GC86	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
SBR2046	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
RC25	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC19	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATGCC
SBR2016	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC7	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC14	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC99	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC11	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC73	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC90	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
[	851				900 ]
SBR1024	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR1015	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
GC86	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR2046	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC25	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC19	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR2016	TGACGCTGAG	ACACGAAAAC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC7	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC14	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC99	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC11	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC73	TGACGCTGAG	ACACGAAAGC	GTGGGGNGCA	AACAGGATTA	GATACCCTGG
RC90	TGACGCTCAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG

*Fig. 8 (continued)*

11/17

[ 901 950 ]

SBR1024	TAGTCCACGC	CTTAAACGAT	GGATACTAAG	TGTCGGCGG	.....
SBR1015	TAGTCCACGC	CTTAAACGAT	GGATACTAAG	TGTCGGCGG	.....
GC86	TAGTCCACGC	CTTAAACGAT	GGATACTAAG	TGTCGGCGG	.....
SBR2046	TAGTCCACGC	CTTAAACGAT	GGATACTAAG	TGTCGGCGG	.....
RC25	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC19	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
SBR2016	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC7	TAGTCCACGC	CCTAAGCTAT	GGATACTAAG	TGTCGGCGG	.....
RC14	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC99	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC11	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC73	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....
RC90	TAGTCCACGC	CCTAAACTAT	GGATACTAAG	TGTCGGCGG	.....

[ 951 1000 ]

SBR1024	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
SBR1015	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
GC86	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
SBR2046	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC25	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC19	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
SBR2016	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC7	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC14	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC99	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC11	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC73	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA
RC90	.....G	TTA.....	.....	.CCGCCGGTG	CCGCAGCTAA

[ 1001 1050 ]

SBR1024	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
SBR1015	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
GC86	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
SBR2046	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC25	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC19	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
SBR2016	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC7	CGCGTTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC14	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC99	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC11	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC73	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA
RC90	CGCATTAAGT	ATCCCGCCTG	GGAAGTACGG	CCGCAAGGTT	GAAACTCAAA

Fig. 8 (continued)



12/17

```

[ 1051                                     1100 ]
SBR1024GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR1015GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
GC86 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR2046GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC25 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC19 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR2016GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCTTGTGGT TTAATTCGAC
RC7 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC14 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC99 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC11 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
RC73 GGGATTGACG GGGGCCCCGCA CAAGCGGTGG GGCATGTGGT TTAATTCGAC
RC90 GGAATTGACG GGGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC

[ 1101                                     1150 ]
SBR1024GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
SBR1015GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
GC86 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
SBR2046GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC25 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC19 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
SBR2016GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC7 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC14 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC99 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC11 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC73 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG
RC90 GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG..... ...CAGGTAG

[ 1151                                     1200 ]
SBR1024TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
SBR1015TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
GC86 TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
SBR2046TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
RC25 TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC19 TAGAAGGGT. .GAAA..GNC TAACGAGGTA .....GCAA. ....TACCAT
SBR2016TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC7 TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC14 TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC99 TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC11 TANAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC73 TNGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT
RC90 TAGAAGGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAT

```

*Fig. 8 (continued)*

13/17

```

[ 1201
SBR1024CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG 1250 ]
SBR1015CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
GC86 CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
SBR2046CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC25 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC19 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
SBR2016CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC7 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC14 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC99 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC11 CGTGCTCAGG TGCTGCATGG CTGTCTTCAG CTCGTGCCGT GAGGTGTTGG
RC73 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
RC90 CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG

[ 1251
SBR1024GTAAAGTCCC GCAACGAGCG CAACCCCTGT CTTAGTTAC CAACGG.... 1300 ]
SBR1015GTAAAGTCCC GCAACGAGCG CAACCCCTGT CTTAGTTAC CAACGG....
GC86 GTAAAGTCCC GCAACGAGCG CAACCCCTGT CTTAGTTAC CAACGG....
SBR2046GTAAAGTCCC GCAACGAGCG CAACCCCTGT CTTAGTTAC CAACGG....
RC25 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC19 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
SBR2016GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC7 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC14 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC99 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC11 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC73 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
RC90 GTAAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TGCCGG....

[ 1301
SBR1024GTCATG.... CCGGGAAGTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG 1350 ]
SBR1015GTCATG.... CCGGGAAGTC TGGAGAGACT GCCCAGGAGA ACGGGGGAGG
GC86 GTCATG.... CCGGGAAGTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG
SBR2046GTCATG.... CCGGGAAGTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG
RC25 GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
RC19 GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
SBR2016GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
RC7 GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGGGGAGG
RC14 GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
RC99 GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
RC11 GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
RC73 GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
RC90 GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG

```

Fig. 8 (continued)

14/17

```

[ 1351                                     1400 ]
SBR1024AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
SBR1015AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
GC86 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
SBR2046AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC25 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC19 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
SBR2016AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC7 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC14 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC99 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC11 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC73 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC
RC90 AAGGTGGGGA TGACGTCAAG TCAGCATGGC CTTTATGCCT GGGGCCACAC

[ 1401                                     1450 ]
SBR1024ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT AAGGGGGAGC
SBR1015ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT AAGGGGGAGC
GC86 ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT AAGGGGGAGC
SBR2046ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT AAGGGGGAGC
RC25 ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT GAGGGGGAGC
RC19 ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT GAGGGGGAGC
SBR2016ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT GAGGGGGAGC
RC7 ACGTGCTACA ATGGCCGGTA CAAAACGCTG CAAACCC.GT GAGGGGGAGC
RC14 ACGTGCTACA ATGGCCGGTA TAAAACGCTG CAAACCC.GT GAGGGGGAGC
RC99 ACGTGCTACA ATGGCCGGTA CAAAACGCTG CAAACCC.GT GAGGGGGAGC
RC11 ACGTGCTACA ATGGCCGGTA CAAAGCGCTG CAAACCC.GT GAGGGGGAGC
RC73 ACGTGCTACA ATGGCCGGTA CAAAACGCTG CAAACCC.GT GAGGGGGAGC
RC90 ACGTGCTACA ATGGCCGGTA CAAAACGCTG CAAACCC.GT GAGGGGGAGC

[ 1451                                     1500 ]
SBR1024CAATCCCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
SBR1015CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
GC86 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
SBR2046CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC25 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC19 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
SBR2016CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC7 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC14 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC99 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC11 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC73 CAATCGCAAA AAACCGGCCT CAGTTCAGAT TGAGGTCTGC AACTCGACCT
RC90 CAATCGCAAA AAACCGGCCT CAGTTCANAT TGAGGTCTGC AACTCGACCT

```

*Fig. 8 (continued)*

15/17

```

[ 1501                                     1550 ]
SBR1024CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
SBR1015CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
GC86 CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
SBR2046CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
RC25 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC19 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
SBR2016CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC7 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC14 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC99 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC11 CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC73 CATGAATGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
RC90 CATGAATGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT

[ 1551                                     1600 ]
SBR1024ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
SBR1015ACGTTCCCGG ACCTTGCTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
GC86 ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
SBR2046ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
RC25 ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC19 ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
SBR2016ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC7 ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC14 ACGTTCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC99 ACGTNCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC11 ACGTNCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC73 ACGTNCCCGG GCCTTGCTACA CACCGCCCGT CACACCACGA AAGCTTGTTG
RC90 ACGTNCCCGG GCCTTGCTACA CCGCGCCCGT CACACCACGA AAGCTTGTTG

[ 1601                                     1650 ]
SBR1024TACCTGAAGT CGTTGGCGCC AACC..... GCAA..... GGAGGCAGAC
SBR1015TACCTGAAGT CGTTGGCGCC AACC..... GCAA..... GGAG-----
GC86 TACCTGAAGT CGTTGGCGCC AACC..... GCAA..... GGGGGCAGAC
SBR2046TACCTGAAGT CGTTGGCGCC AACC..... GCAA..... GGAGGCAGAC
RC25 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC19 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
SBR2016TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC7 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC14 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC99 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GAAGGCAGGC
RC11 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC73 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCAGGC
RC90 TACCTGAAGT CGCCCAAGCC AACC..... GCAA..... GGAGGCANGC

```

Fig. 8 (continued)

16/17

```

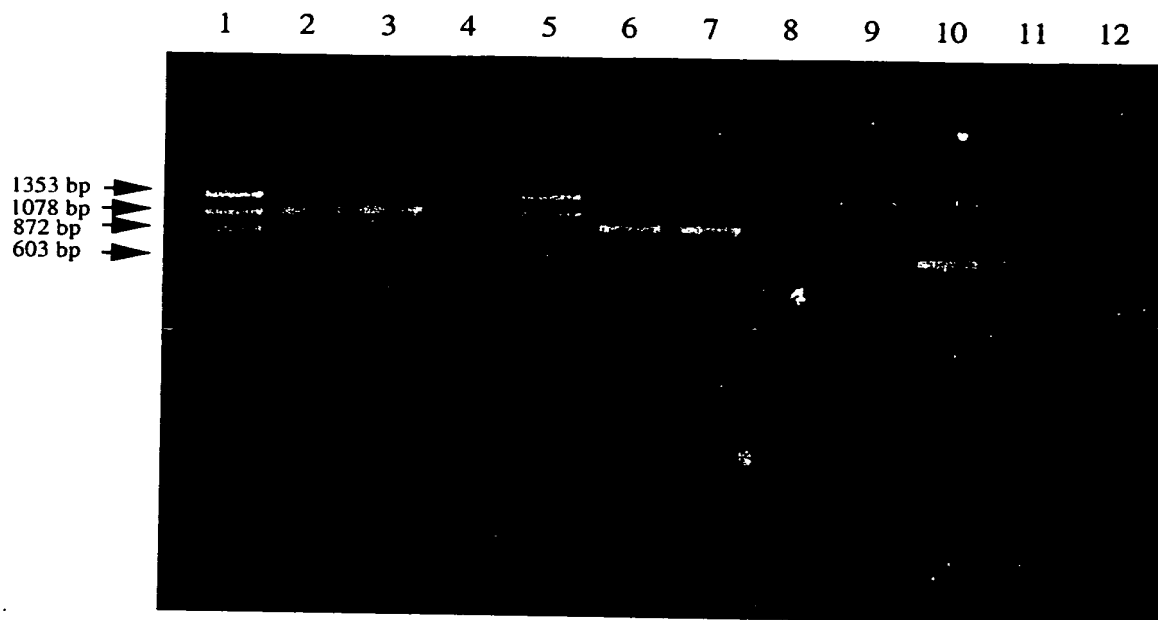
[ 1651                                                                    1700 ]
SBR1024GCCCACGGTA TGACCGATGA TTGGG-----
SBR1015-----
GC86   GCCCACGGTA TGACCGATGA TTGGGGTGAA GTCGTAACAA GGTAACCGTA
SBR2046GCCCACGGTA TGACCGATGA TTGGGG-----
RC25   GCCCACGGTA TGGCCCGTGA TTGGGGTGAA GTCGTAACAA GGTAACCGTA
RC19   GCCCACGGTA TGGCCCGTGA TTGGGGTGAA GTCCTAACA-----
SBR2016GCCCACGGTA TGGC-----
RC7    GCCCACGGTA TGGCCG----
RC14   GCCCACGGTA TGGCCCGTGA T-----
RC99   GCCCACGGTA TGGCCCGTGA -----
RC11   GCCCACGGTA TGGCCCGTGA TGGGG-----
RC73   GCCCACGGTA TGGCCCGTGA TGGGG-----
RC90   GCCCACGGTA TGGCCCGTGA TG....-----

[ 1701                                                                    1750 ]
SBR1024-----
SBR1015-----
GC86   ATC-----
SBR2046-----
RC25   AA-----
RC19   -----
SBR2016-----
RC7    -----
RC14   -----
RC99   -----
RC11   -----
RC73   -----
RC90   -----
;

```

*Fig. 8 (continued)*

17/17



*Fig. 9*

BEST AVAILABLE COPY